Transcriptomic landscape of lncRNAs in inflammatory bowel disease

Mirza, Aashiq Hussain; Bang-Berthelsen, Claus Heiner; Seemann, Ernst Stefan; Pan, Xiaoyong; Frederiksen, Klaus S; Vilien, Mogens; Gorodkin, Jan; Pociot, Flemming

Published in:
Genome Medicine

DOI:
10.1186/s13073-015-0162-2

Publication date:
2015

Document version
Publisher's PDF, also known as Version of record

Genome Medicine

BioMed Central
The Open Access Publisher

Transcriptomic landscape of lncRNAs in inflammatory bowel disease

Genome Medicine (2015) 7:39

doi:10.1186/s13073-015-0162-2

Aashiq H Mirza (mirza@sund.ku.dk)
Claus HB Berthelsen (clbb@novonordisk.com)
Stefan E Seemann (seemann@rth.dk)
Xiaoyong Pan (panxy@rth.dk)
Klaus S Frederiksen (ksf@novonordisk.com)
Mogens Vilien (mogens.vilien@regionh.dk)
Jan Gorodkin (gorodkin@rth.dk)
Flemming Pociot (flemming.pociot.01@regionh.dk)

Published online: 13 May 2015

ISSN 1756-994X

Article type Research

Submission date 28 November 2014

Acceptance date 9 April 2015

Article URL http://dx.doi.org/10.1186/s13073-015-0162-2

Transcriptomic landscape of lncRNAs in inflammatory bowel disease

Aashiq H Mirza[1,2,3,†]
Email: mirza@sund.ku.dk

Claus HB Berthelsen[1,6,†]
Email: clbb@novonordisk.com

Stefan E Seemann[1,3]
Email: seemann@rth.dk

Xiaoyong Pan[1,3,5]
Email: panxy@rth.dk

Klaus S Frederiksen[7]
Email: ksf@novonordisk.com

Mogens Vilien[4]
Email: mogens.vilien@regionh.dk

Jan Gorodkin[1,3]
Email: gorodkin@rth.dk

Flemming Pociot[1,2,3,*]
[*] Corresponding author
Email: flemming.pociot.01@regionh.dk

[1] Center for non-coding RNA in Technology and Health, University of Copenhagen, Copenhagen, Denmark

[2] Department of Pediatrics E, Copenhagen Diabetes Research Center (CPH-DIRECT), Herlev University Hospital, Herlev 2730, Denmark

[3] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[4] Department of Surgery, North Zealand Hospital, Hillerød 3400, Denmark

[5] The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

[6] Department of Obesity Biology, Novo Nordisk, Måløv, Denmark

[7] Department of Molecular Genetics, Novo Nordisk, Måløv, Denmark

[†] Equal contributors.

Abstract

Background

Inflammatory bowel disease (IBD) is a complex multi-factorial inflammatory disease with Crohn's disease (CD) and ulcerative colitis (UC) being the two most common forms. A number of transcriptional profiling studies have provided compelling evidence that describe the role of protein-coding genes and microRNAs in modulating the immune responses in IBD.

Methods

In the present study, we performed a genome-wide transcriptome profiling of lncRNAs and protein-coding genes in 96 colon pinch biopsies (inflamed and non-inflamed) extracted from multiple colonic locations from 45 patients (CD = 13, UC = 20, Controls = 12) using expression microarrays platform.

Results

In our study, we identified widespread dysregulation of lncRNAs and protein-coding genes in both inflamed and non-inflamed CD and UC compared to the healthy controls. In case of inflamed CD and UC, we identified 438 and 745 differentially expressed lncRNAs, respectively, while in case of the non-inflamed CD and UC, we identified 12 and 19 differentially expressed lncRNAs, respectively. We also observed significant enrichment (p-value < 0.001, Pearson's Chi-squared test) for 96 differentially expressed lncRNAs and 154 protein-coding genes within the IBD susceptibility loci. Furthermore, we found strong positive expression correlations for the intersecting and *cis*-neighboring differentially expressed IBD loci-associated lncRNA-protein-coding gene pairs. The functional annotation analysis of differentially expressed genes revealed their involvement in immune response, pro-inflammatory cytokine activity and MHC protein complex.

Conclusions

The lncRNA expression profiling in both inflamed and non-inflamed CD and UC, successfully stratified IBD patients from the healthy controls. Taken together, the identified lncRNA transcriptional signature along with clinically relevant parameters suggests their potential as biomarkers in IBD.

Background

Inflammatory bowel diseases (IBD) are idiopathic chronic relapsing inflammatory conditions of the gastrointestinal (GI) tract. Crohn's disease (CD) and ulcerative colitis (UC) are two most common forms of the IBD. IBD is emerging as a global disease with its incidence and prevalence differentially increasing geographically around the world. Accumulating evidence suggests that IBD result from the complex interplay between genetic, immunologic, and modifiable environmental factors [1], in a genetically susceptible host against a subset of gut commensal microbiota [2-4].

CD is characterized by intestinal inflammation in a discontinuous fashion and involves any part of the GI tract, although in majority of the cases terminal ileum and/or colon is affected. Transmural pattern of inflammation is a hallmark of CD accompanied with other pathophysiological complications like thickened submucosa, intestinal fibrosis, fissuring ulceration in highly active disease, non-caseating granulomas, strictures, abscesses and fistulas [3]. By contrast, UC involves only rectum and colon, and is characterized by superficial inflammation that is restricted to the mucosa and submucosa with presence of cryptitis and crypt abscesses. Disease activity in both CD and UC is typically relapsing and remitting and both conditions are often difficult to diagnose because of idiosyncrasies in the presentation of overlapping and distinct clinical and pathological features [2,3]. Characteristically, diagnosis of either CD or UC is based on number of findings including: clinical symptoms, endoscopic features, radiologic tests, and biopsy histology.

According to a recent meta-analysis of IBD genome-wide association studies (GWAS) data, the number of confirmed genetic loci associated with risk for IBD has increased to 163, with 110 shared between CD and UC, 30 CD-specific and 23 UC-specific. Interestingly, an overwhelming majority of these IBD loci are located in the noncoding intergenic and intronic regions [5]. Most are overlapping the regulatory elements and consequently are likely to influence gene regulation. Findings from our recent studies have demonstrated that large number of annotated long non-coding RNAs (lncRNAs), including novel evolutionarily conserved structured RNA motifs with regulatory potential ([6], unpublished observations (Seemann et al.)), overlap with the IBD loci. Consistent with our findings, another recent study, elegantly revealed that IBD loci overlaps with active regulatory regions in primary intestinal epithelium and immune cells and also IBD loci was found significantly enriched within these active regulatory regions [7].

Several transcriptome profiling studies have provided compelling evidence describing the role of protein-coding and non-coding RNAs (ncRNAs), such as microRNAs, in modulating immune responses in IBD [8-15]. In murine models, loss of endogenous intestinal microRNAs is known to cause impairment of epithelial barrier function that result in acute inflammation [16]. Several studies have explored clinical differences between CD and UC based on transcriptional regulation [17,18]. Recently, Granlund et al. demonstrated lack of major differences between CD and UC based on protein-coding gene expression profiling in IBD [9]. In contrast, expression profiling of colon biopsies from IBD patients allude to differential diagnosis of CD and UC based on transcriptional signatures associated with intestinal inflammation [19].

Long non-coding RNAs (lncRNAs) have emerged as important regulators of gene expression, with accumulating body of evidence linking lncRNAs to a plethora of human pathologies including inflammatory diseases [20]. Although, the precise role of lncRNAs in intestinal diseases remains poorly understood, evidence from recent studies indicates that lncRNAs might be playing a crucial role in inflammatory cascades. Indeed, preponderance of emerging evidence from number of studies, demonstrates important roles of lncRNAs in regulating gene expression within the immune system. Nevertheless, identification of IBD susceptibility loci has afforded limited success in translating results from gene expression studies to advance our knowledge and understanding of the IBD pathophysiology. In particular, the details about the initiation, propagation and maintenance of the lingering inflammation in IBD remains unclear. Furthermore, earlier transcriptomic studies in IBD have mostly been focused on the protein-coding genes with limited profiling studies focusing on microRNAs. However, no study has explored the genome-wide expression profile of lncRNAs in IBD.

In the present study, transcriptomic profiling of lncRNAs and protein-coding genes from colon pinch biopsies of IBD patients was performed using expression microarray platform. Our results identified widespread dysregulation of lncRNAs and protein-coding gene expression in both CD and UC. Notably, differential transcriptomic signature of lncRNAs and protein-coding genes in inflamed-CD (iCD) and inflamed-UC (iUC) enabled clear stratification of the CD and UC phenotype. These data indicate that lncRNAs could potentially be used as predictive biomarkers in IBD.

Methods

Samples collection (Patients and control)

All the patient samples were collected from an IBD cohort at North Zealand Hospital, Hillerød, Denmark. Subjects were required to meet the Copenhagen criteria for Crohn's disease or ulcerative colitis. Participants recruited for the study were patients admitted to the Department of Gastroenterology for colonoscopy who were diagnosed either with CD or UC, or were admitted to the clinic for diagnostic colonoscopy because of symptoms unrelated to the IBD. Written informed consent from all the participants in the study was acquired prior to the collection of samples and medical history. In total, 90 biopsies were collected from 45 individuals (13 CD, 20 UC patients and 12 healthy controls). Subjects were included as normal controls only after all clinical examinations had concluded no signs of autoimmune or inflammatory disease. In case of the IBD groups (CD and UC), one to five endoscopic pinch biopsies were extracted from macroscopically most inflamed mucosa (iCD/iUC) and adjacent non-inflamed (niCD/niUC) mucosa within colon (transverse, descending, sigmoid), ileum, transverse ileum and rectum for the CD, and colon sigmoid and rectum for the UC patients. Whereas, for the control group, one to five biopsies were taken from the same locations as in the CD group, except for the two samples (G3_G1 and 60_G1) that were extracted from duodenal bulb. All biopsies were placed in RNA*later* solution (QIAGEN, Hilden, Germany), and stored for the later downstream use. The study was approved by the Regional Ethical Committee (H-4-2012-030). The inflammation status of biopsies was confirmed by the histologic examination and features of chronic intestinal inflammation for each patient were scored using a previously described scoring system for UC [21] and CD [22]. The pathologists were blinded to the status of inflammation. Additionally, we also tested expression of a panel of twenty-six pro-inflammatory markers (cytokines, interleukins, metalloproteases) using qPCR (Fluidigm platform) to confirm the inflammation status of biopsies (data not shown) prior to the microarray analysis.

RNA extraction and quality control

Total RNA was extracted from biopsies stored in RNA*later* using RNeasy Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Briefly, the biopsy samples were homogenized in lysis buffer with 1.4 mm ceramic beads (MO BIO Laboratories) using Thermo Savant FastPrep FP120 Homogenizer for 30 seconds at a speed of 4 m/sec. All the remaining steps of the protocol were performed according to the manufacturer's recommendations. To remove traces of genomic DNA, samples were treated with DNase I (QIAGEN, Hilden, Germany). RNA was finally eluted with nuclease-free water supplied with the kit. The quantity and purity of isolated RNA was determined by UV absorbance using NanoDrop 2000 Spectrophotometer (Thermo Scientific, DE, USA), and the integrity of RNA was assessed by analysis of rRNA band integrity on an Agilent 2100

Bioanalyzer RNA 6000 LabChip kit (Agilent Technologies). Only RNA samples with RNA integrity number (RIN) >7 were used for the microarray experiments.

Microarray hybridization

100 ng of total RNA was labeled using LowInputQuick Amp Labeling kit v6.5 (Agilent 5190–2305) following manufacturer instructions. Briefly, mRNA was reverse transcribed in the presence of T7-oligo-dT primer to synthesize cDNA. The cDNA was then in-vitro transcribed with T7 RNA polymerase in the presence of Cy3-CTP to generate labeled cRNA. The labeled cRNA was hybridized to the Agilent Custom 8x60K format lncRNA expression microarray (AMADID 047718, based on Gencode v15 catalog of human long noncoding RNAs, probe length = 60 nt) according to the manufacturer's protocol. Finally, the arrays were washed, and scanned on an Agilent G2565CA microarray scanner at 100% PMT and 3 μm resolution. Intensity data was extracted using the Feature Extraction software (Agilent). A more detailed and general information about the array can also be found on GENCODE website [23]. The raw microarray data reported in this manuscript have been deposited in the Gene Expression Omnibus (GEO) database with the accession number GSE67106.

Statistical analyses

Raw data was corrected for background noise using normexp method [24]. To assure comparability across samples we used quantile normalization [unpublished observations (Bolstad B. Probe Level Quantile Normalization of High Density Oligonucleotide Array Data. 2001. [25]). Median intensity was taken between technical replicates after checking pairwise Pearson correlation coefficients ($r^2 \geq 0.98$). Differential expression analysis was carried out on non-control probes with an empirical Bayes approach on linear models (LIMMA) [26]. Principal Component Analysis (PCA) method was employed for the initial interpretation of the data. In total, we made seven contrasts to identify differentially expressed genes (iCD vs control, iUC vs control, iCD vs niCD, iUC vs niUC, niCD vs control, niUC vs control and iCD vs iUC) (Additional file 1: Table S1, S2, S3, S4).. P-values were adjusted for multiple comparisons using the False Discovery Rate (FDR) correction [27]. Differentially expressed genes were identified using the double-filtering criterion: adjusted p-value (FDR) < 0.05 and an absolute fold change (abs FC) > 1.5. For transcripts targeted by two probes, only those probes that were changing in the same direction and the probes with highest FC values were retained for further analysis. All statistical analyses were performed with Bioconductor in R statistical environment [28].

Validation of differentially expressed genes by quantitative real-time PCR (qPCR)

The expression of differentially expressed genes from microarray experiments was validated by the quantitative real-time PCR (qPCR) using hydrolysis probe based inventoried and custom designed PrimeTime qPCR 5′ Nuclease assays procured from Integrated DNA Technologies (IDT). The Double-Quenched hydrolysis probes with 5′ FAM fluorophore, a 3′ IBFQ quencher, and an internal ZEN™ quencher was used for all the assays.. From the list of top differentially expressed genes from different contrasts, 6 up- and 6 down-regulated genes were selected for their expression validation by qPCR in a subset of samples used for the microarray experiments. In case of protein-coding, 3 up-regulated (*DUOXA2*, *CHI3L1*, *DST*), and 3 down-regulated (*PCK1*, *KCNK10*, and *SERPINB3*) genes were validated. Whereas, for

the lncRNAs, 3 up-regulated (*MMP12, RP11-731 F5.2*, *AC007182.6*), and 3 down-regulated (*DPP10-AS1*, *CDKN2B-AS1*, and *AL928742.12*) genes were validated. In addition, expression of protein-coding gene *DUOX2* was also measured by qPCR. All cDNAs were prepared using 750 ng of DNA-free RNA using iScript^{TM} cDNA synthesis kit (BioRad) with a mixture of random and oligo(dT) primers following the manufacturer's instructions. Real-time PCR was performed with 7.5 ng of cDNA per well template for the all the protein-coding genes and lncRNAs with Brilliant III Ultra-Fast QPCR Master Mix (Agilent Technologies). For PCR amplification, the following thermal profile was used: 3 min 95°C; 40 x (5 sec 95°C, 10 sec 60°C). Expression of each lncRNA and protein-coding gene tested was represented as a fold change using the $2^{-\Delta\Delta CT}$ method. *GAPDH* was used as the reference gene.

Identification of IBD-loci associated lncRNAs

All IBD loci marker SNPs and associated genes were retrieved from the ImmunoBase [29]. In total, 233 unique marker SNPs for IBD, CD, and UC regions were retrieved and mapped to the 22,007 lncRNAs (Gencode v15) using intersect feature of the BedTools [30]. The susceptibility locus for IBD was defined based on 500 kb long genomic region with the IBD marker SNP in the middle. The differentially expressed lncRNAs from five contrasts (iCD vs control, iUC vs control, iCD vs niCD, iUC vs niUC, and iCD vs iUC), were mapped to the IBD loci to identify the IBD loci-associated lncRNAs. Regulatory evidence for the IBD-associated SNPs was retrieved from Mokry et al. [7] and RegulomeDB [31].

Functional annotation and Gene Ontology (GO) analysis of differentially expressed lncRNAs

For the differentially expressed lncRNAs, the nearest protein-coding neighbors within a span of <10 kb were identified. For the antisense overlapping or intronic overlapping lncRNAs, intersecting protein-coding gene(s) were identified using intersect feature of BedTools [30]. PANTHER (protein annotation through evolutionary relationship) classification system [32] was used to perform functional annotation and GO analysis of genes that overlap with or are neighbors of the differentially expressed lncRNAs. Likewise, for the IBD loci-associated lncRNAs, GO analysis was performed using the above described nearest neighbor approach. The enrichment for over-represented GO functional terms was calculated based on Binomial test in PANTHER.

Sample classification using SVM based on differentially expressed genes identified by LIMMA

Support Vector Machines (SVM) [33] was used for classifying the CD and UC cases from the controls based on differentially expressed genes identified by LIMMA in five contrasts (iCD vs control, iCD vs niCD, iUC vs control, iUC vs niUC and iCD vs iUC). SVM classification was applied to all five contrasts using the Leave-one-out cross-validation (LOOCV) for differentially expressed lncRNAs and protein-coding genes. To explore the effect of various clinical parameters (age, sex, smoking, disease index and biopsy location) on overall disease outcome, we used the following linear regression function:

$$y = err + w1 * age + w2 * sex + w3 * smoking + w4 * disease\ index + w5^{*} biopsy\ location$$

Here, y = 1 for iCD or iUC disease phenotype and for rest of the samples it is 0. For clinical parameters: age and disease index, we used original values, while for sex and smoking, we used the following binary outcomes: male = 1, female = 0 and smoker =1, non-smoker = 0. In case of six biopsy locations, we used values ranging from 0 to 5. To control any input bias, same analysis was performed on a randomized lncRNA gene list of the same number as the total differentially expressed lncRNA genes. The feature values were normalized to values ranging from 0 to 1 using (x-min)/(max-min). Linear regression was applied using Scikit-learn [34] package in python and least square method was used for optimization in our analysis. Furthermore, differentially expressed genes identified by LIMMA were verified by Support Vector Machines- Recursive Feature Elimination (SVM-RFE) method [35]. SVM-RFE recursively prunes genes whose absolute weights are the smallest until desired number of features is reached. For each contrast, we used SVM-RFE to identify the same number of differentially expressed genes as identified by LIMMA.

Co-expression network analysis

To identify CD and UC specific networks clusters (modules) based on highly correlated genes, weighted correlation network analysis (WGCNA) method was used [36]. We used the normalized expression data as input and removed the outlier samples. The clinical parameters were represented as following: numeric for age and disease, binary for sex, ethnicity (3 categories), smoking (4 categories), clinical subgroup (5 categories), and biopsy location (6 categories). The standard procedure of WGCNA was applied for network construction and module identification. The trait-based gene significance measure is defined as the absolute correlation and correlation test p-value between the trait and the gene expression profile. Gene Ontology analysis of modules was performed with GOstats package in R [37] using adjusted p-value < 0.001. We controlled for study bias in the GO analysis by running the same analysis for randomized gene sets of the same module sizes.

Results

The overall summary describing sample information is provided in Table 1. Both, CD and UC samples were divided based on inflammation status confirmed by both macroscopic, microscopic evaluations and pro-inflammatory gene signatures into inflamed (iCD, iUC) and non-inflamed (niCD, niUC) categories. The total number of samples included 21 iCD, 23 niCD, 15 iUC, 9 niUC and 22 healthy controls. In total, 90 intestinal pinch biopsies (45 individuals (13 CD, 20 UC patients and 12 healthy controls)) from multiple colonic regions were harvested from both inflamed and non-inflamed mucosa (Figure 1A). Detailed sample information including the ethnic background, disease index, previous treatment regimens and other clinical parameters are listed in Table 2.

Table 1 Overall study design and sample information

Diagnosis	Number of samples	Number of individuals
iCD	21	13
niCD	23	
iUC	21 (15 unique)	20
niUC	9	
Controls	22	12
Total	96 (90 unique samples)	45

90 biopsy samples extracted from different colonic locations from 45 patients (CD = 13, UC = 20, Controls = 12). 6 samples from UC patients were used as technical replicates.

Figure 1 Study design and inflammatory gene signature for iCD and iUC. **(A)** Study design included 90 pinch biopsies from multiple colonic regions for both inflamed and non-inflamed mucosa (21 iCD, 23 niCD, 15 iUC, 9 niUC and 22 healthy controls samples). **(B)** Principal component analysis (PCA) separated inflamed-CD (iCD) and inflamed-UC (iUC) samples from non-inflamed and healthy controls. PC1 and PC2 together explained 15% of the total variation. **(C)** Unsupervised hierarchical clustering of the most dynamic probes (coefficient of variance >0.05) across the samples resulted in clustering of samples according to their clinical subgroups. **(D)** The log2 ratio and –log10 adj.P-values are plotted in the form of volcano plots for iCD vs control, iUC vs control and iCD vs iUC contrasts. The probes in red, blue and orange colors represents up-regulated (FC >1.5 and adj.P-value <0.05), down-regulated (FC < −1.5 and adj.P-value <0.05) and significant with small fold change (FC > −1.5 and <1.5), respectively. The non-significant probes are represented in black color. The selected protein-coding genes and lncRNAs labeled in black and green, respectively. **(E)** Venn diagram shows the overlap between differentially expressed genes identified in iCD vs control, iUC vs control and iCD vs iUC contrasts. The up-regulated genes are depicted in italics, down-regulated as underlined and contra-regulated in red.

Table 2 Clinical parameters

Number of individuals	**CD** **13**	**UC** **20**	**Controls** **12**
Age median years (range)	31 (19–59)	46 (18–68)	54 (18–77)
Average age at diagnosis (years)	27	33	NA
Average years with disease (Disease duration)	8	9.3	NA
Female / Male	6/7	13/7	8/4
Smoking			
Smoker (S)	8	1	4
Previous (P)	4	8	1
Never (N)	1	11	5
Not disclosed (ND)	-	-	2
Ethnicity			
Danish (DK)	9	19	12
European (EU)	1	-	-
Middle Eastern (ME)	3	1	-
Number of individuals with family history of other autoimmune diseases	1 (7%)	6 (30%)	5 (41%)
Number of patients on medication			NA
5-ASA	2	13	
Solumedrol	2	1	
Azathioprin	2	2	
Budesonide	1	1	
Prednisolon	1	2	
Disease Index	HB index = 3-36	SCCAI index = 2-12	

Each column summarizes characteristics for all patients contributing with samples to the corresponding sample groups. 5-ASA – 5-aminosalicylic acid. HB index – Harvey Bradshaw index. SCCAI index – Simple Clinical Colitis Activity index.

Microarray analysis of lncRNAs and protein-coding gene expression

In Gencode v15 lncRNA microarray design, each lncRNA transcript is targeted by two probes covering 22,001 lncRNA transcripts corresponding to 12,963 lncRNA genes. In addition, each array contains 17,535 randomly-selected protein-coding targets, of which 15,182 (unique 12,787) correspond to protein-coding genes. Six samples analyzed in duplicates, hybridized on separate chips, and used as technical replicates showed strong positive Pearson correlation ($r^2 \geq 0.98$, p-value < 2.2e-16) (Additional file 2: Figure S1). Based on the principal component analysis (PCA) (see methods section for details), separation of iCD and iUC samples from niCD, niUC and healthy controls were observed (Figure 1B). However, there was no apparent separation between iCD and iUC samples. The scatterplot matrices describing the first four principal components are described in Additional file 2: Figure S2. Unsupervised hierarchical clustering of the most dynamic probes (coefficient of variance >0.05) across the samples ensued in clustering of samples according to their clinical subgroups (Figure 1C). The probes targeting lncRNAs and protein-coding genes separately also clustered samples in a similar manner (Additional file 2: Figure S3A and S3B).

Differential transcriptional signature of lncRNAs and protein-coding genes in CD and UC

To define CD and UC specific transcriptional signatures based on intestinal inflammation, we identified differentially expressed genes using LIMMA [24] (based on a cutoff of log2 fold change (FC) >1.5 (up-regulated), FC < −1.5 (down-regulated) and adj.P-value <0.05 (moderated t-test)) in all contrasts (Table 3). The log2 ratio and –log10 adj.P-values are plotted and represented as volcano plots for iCD vs control, iUC vs control and iCD vs iUC contrasts in Figure 1D. For the non-inflamed tissues contrasts (iCD vs niCD and iUC vs niUC), the volcano plots are shown in Additional file 2: Figure S4A and S4B, respectively.

Table 3 Total number of differentially expressed genes

Total differentially expressed genes				
	iCD	**niCD**	**iUC**	**niUC**
Control	1477	73	2429	44
iCD		435	73	
iUC				1814
Protein-coding genes				
	iCD	**niCD**	**iUC**	**niUC**
Control	1039	61	1684	25
iCD		328	50	
iUC				1215
LncRNAs				
	iCD	**niCD**	**iUC**	**niUC**
Control	438	12	745	19
iCD		107	23	
iUC				599

Total differentially expressed genes identified in seven pairwise contrasts (iCD vs control, iUC vs control, iCD vs niCD, iUC vs niUC, niCD vs control, niUC vs control and iCD vs iUC).

Differential gene expression analysis identified the following up/down-regulated genes: 761/278 protein-coding genes and 254/184 lncRNAs in iCD vs control and 1085/599 protein-coding genes and 370/375 lncRNAs in iUC vs control (Table 3 and Figure 1E). The top up-regulated and down-regulated lncRNAs and protein-coding genes (based on FC) for iCD vs control and iUC vs control are listed in Tables 4 and 5. Interestingly, lncRNA *RP11-731 F5.2* (whose 3′ end partly spans into the start of the *IGHG2* gene) and antisense lncRNA *MMP12* were found significantly up-regulated, whereas, the antisense *DPP10-AS1, ANRIL* (*CDKN2B-AS1)* and *DIO3OS* lncRNAs were significantly down-regulated in both iCD vs control and iUC vs control contrasts (Tables 4 and 5).

Table 4 Top ten differentially expressed genes in iCD

Up-regulated lncRNAs			**Up-regulated protein-coding genes**		
Gene Name	**Transcript**	**FC**	**Gene Name**	**Transcript**	**FC**
RP11-731 F5.2	ENST00000460164.1	14.14	REG3A	NM_138938	52.71
MMP12	ENST00000532855.1	6.64	DUOXA2	NM_207581	47.26
MMP12	ENST00000326227.5	6.52	DEFA5	NM_021010	37.73
RP11-465 L10.10	ENST00000419897.1	5.69	DEFA6	NM_001926	28.33
RP11-44 K6.2	ENST00000520185.1	3.83	CHI3L1	NM_001276	26.29
FAM66D	ENST00000526690.1	3.36	CXCL1	NM_001511	14.8
LINC01272	ENST00000445003.1	3.35	DMBT1	NM_007329	13.45
RP11-44 K6.4	ENST00000522970.1	3.24	SAA1	NM_000331	12.67
SAA2-SAA4	ENST00000524555.1	3.16	CXCL9	NM_002416	12.07
KIF9-AS1	ENST00000429315.2	3.14	IGHG3	ENST00000390551	11.52
Down-regulated lncRNAs			**Down-regulated protein-coding genes**		
DPP10-AS1	ENST00000432658.1	−8.57	PCK1	NM_002591	−5.55
PDZK1P2	ENST00000401008.2	−4.11	SLC26A2	NM_000112	−3.82
DIO3OS	ENST00000553575.1	−3.01	C10orf116	NM_006829	−3.8
DIO3OS	ENST00000554694.1	−3.01	GUCA2B	NM_007102	−3.61
DIO3OS	ENST00000557532.1	−2.99	LCN15	NM_203347	−3.43
DIO3OS	ENST00000557109.1	−2.98	AQP7P1	NR_002817	−3.32
ANRIL (CDKN2B-AS1)	ENST00000422420.1	−2.97	TRPM6	NM_017662	−3.19
ANRIL (CDKN2B-AS1)	ENST00000428597.1	−2.97	TNNC2	NM_003279	−3.1
DIO3OS	ENST00000554441.1	−2.96	UGT2A3	NM_024743	−2.97
DIO3OS	ENST00000554735.1	−2.95	ADH1C	NM_000669	−2.96

Top ten up and down-regulated lncRNAs and protein-coding genes in iCD vs control contrast. The log2 fold change is denoted as FC.

Table 5 Top ten differentially expressed genes in iUC

Up-regulated lncRNAs			**Up-regulated protein-coding genes**		
Gene Name	**Transcript**	**FC**	**Gene Name**	**Transcript**	**FC**
RP11-731 F5.2	ENST00000460164.1	20.64	DUOXA2	NM_207581	109.61
MMP12	ENST00000532855.1	17.05	CHI3L1	NM_001276	39.71
MMP12	ENST00000326227.5	16.54	SAA1	NM_000331	30.67
RP11-465 L10.10	ENST00000419897.1	9.52	CXCL1	NM_001511	25.92
KIF9-AS1	ENST00000429315.2	5.75	MMP7	NM_002423	21.2
FAM66D	ENST00000526690.1	5.73	SLC6A14	NM_007231	20.52
SAA2-SAA4	ENST00000524555.1	5.66	IGHG3	ENST00000390551	20.14
CLRN1-AS1	ENST00000476886.1	4.64	MMP12	NM_002426	15.76
RP11-1149O23.3	ENST00000517774.1	4.29	C4orf7	NM_152997	14.76
RP5-1028 K7.2	ENST00000578280.1	4.21	CXCL2	NM_002089	11.91
Down-regulated lncRNAs			**Down-regulated protein-coding genes**		
ANRIL (CDKN2B-AS1)	ENST00000422420.1	−8.67	PCK1	NM_002591	−15.24
ANRIL (CDKN2B-AS1)	ENST00000428597.1	−8.31	OSTalpha	NM_152672	−11.33
ANRIL (CDKN2B-AS1)	ENST00000585267.1	−7.06	ANPEP	NM_001150	−11.02
ANRIL (CDKN2B-AS1)	ENST00000580576.1	−6.92	SLC26A2	NM_000112	−10.46
ANRIL (CDKN2B-AS1)	ENST00000577551.1	−6.74	GBA3	NM_020973	−9.28

ANRIL (CDKN2B-AS1)	ENST00000581051.1	−6.72	GUCA2A	NM_033553	−9.22
ANRIL (CDKN2B-AS1)	ENST00000582072.1	−6.68	SLC3A1	NM_000341	−9.21
PDZK1P2	ENST00000401008.2	−6.67	GUCA2B	NM_007102	−8.84
DPP10-AS1	ENST00000432658.1	−5.95	TMIGD1	NM_206832	−8.17
ANRIL (CDKN2B-AS1)	ENST00000421632.1	−5.78	SLC1A7	NM_006671	−6.57

Top ten up and down-regulated lncRNAs and protein-coding genes in iUC vs control contrast. The log2 fold change is denoted as FC.

In case of the protein-coding genes, the top differentially expressed genes included, *DUOXA2*, *CHI3L1*, *CXCL1* and *SAA1*, which were all significantly up-regulated, whereas, *PCK1*, *SLC26A2*, *GUCA2B*, were significantly down-regulated (Tables 4 and 5). In case of iCD vs controls, *REG3A* was >52 fold up-regulated (adj.P-value = 2.17e-04). The top differentially expressed lncRNAs and protein-coding genes for iCD vs niCD and iUC vs niUC contrasts, displayed similar expression patters as healthy controls (Additional file 1: Table S1 and S2).

On comparing niCD vs control and niUC vs control, only a small number of up/down-regulated genes (61/25 and 8/17 protein-coding, 12/19 and 9/10 lncRNAs) were identified for niCD and niUC, respectively. Nearly, all of the differentially expressed genes in niCD vs control were also present in iCD vs control contrast with the exception of protein-coding gene *CRYBB2* (FC = −1.5) (Additional file 1: Table S3). Conversely, in case of niUC vs control, majority (15 out of 17) of the up-regulated genes including 4 small nucleolar RNAs (snoRNAs: *SNORD97*, *SNORA28*, *SNORA53*, and *SNORA74A*) and the down-regulated genes, *MAST3*, *CPT1B*, *LOC338799*, *EXOC3L4*, and *MAPK8IP3* were specifically found in niUC only (Additional file 1: Table S4). Importantly, in the case of iCD vs iUC contrast, 18/32 protein-coding genes and 13/10 lncRNAs were significantly found to be up/down-regulated. The top up/down-regulated lncRNAs and protein-coding genes for iCD vs iUC are shown in Table 6. The annotations for the Gencode v15 [38] microarray features for lncRNAs are summarized in Figure 2A. Majority of the differentially expressed lncRNAs identified in our analysis belonged to three main classes: antisense, processed transcripts and intergenic lincRNAs (Figure 2B), as described in the following section.

Table 6 Top ten differentially expressed genes in iCD vs iUC

Up-regulated lncRNAs			**Up-regulated protein-coding genes**		
Gene Name	**Transcript**	**FC**	**Gene Name**	**Transcript**	**FC**
FLJ42969	ENST00000514926.1	2.6	C8G	NM_000606	2.75
AC007182.6	ENST00000455232.1	2.42	SLC25A34	NM_207348	2.43
RP11-542 M13.2	ENST00000599411.1	2.04	UGT1A6	NM_001072	2.25
RP11-399 F4.4	ENST00000453998.1	1.87	LRRC66	NM_001024611	2.20
FAM95B1	ENST00000455995.1	1.87	EXOC3L4	NM_001077594	1.95
RP3-395 M20.8	ENST00000432521.2	1.69	ANO7	NM_001001891	1.95
RP3-395 M20.8	ENST00000448624.2	1.65	GLYCTK	NM_145262	1.90
OPLAH	ENST00000426825.1	1.61	CLEC10A	NM_182906	1.89
OPLAH	ENST00000534424.1	1.61	FAM95B1	NR_026759	1.89
SPPL2B	ENST00000592738.1	1.59	LPIN3	NM_022896	1.87
Down-regulated lncRNAs			**Down-regulated protein-coding genes**		
AL928742.12	ENST00000412518.1	−2.01	SERPINB3	NM_006919	−3.87
RP11-444D3.1	ENST00000540811.1	−1.84	SLC6A14	NM_007231	−3.51
AL928742.12	ENST00000427543.1	−1.8	GAL	NM_015973	−2.50
FAM25D	ENST00000426412.2	−1.69	GJB4	NM_153212	−2.38
RP11-274 N19.2	ENST00000515643.1	−1.64	IGHV1-58	ENST00000390628	−2.29
RP11-838 N2.4	ENST00000579007.1	−1.59	CRYM	NM_001888	−2.28
RP11-279 F6.3	ENST00000558941.1	−1.57	SLC26A4	NM_000441	−2.22
RP11-279 F6.3	ENST00000559212.1	−1.55	DEFB103B	NM_018661	−2.22

LINC00524	ENST00000555860.1	−1.54	LAMC2	NM_005562	−2.20
VAV3-AS1	ENST00000438318.1	−1.52	TUSC3	NM_178234	−2.02

Top ten up and down-regulated lncRNAs and protein-coding genes in iCD vs iUC. The log2 fold change is denoted as FC.

Figure 2 Gencode v15 annotation of the total differentially expressed lncRNAs in IBD and microarray validation by qPCR. **(A)** The Gencode v15 array targeted 22007 lncRNA transcripts falling into seven major annotation classes (antisense, processed transcripts, intergenic (lincRNAs), sense overlapping, sense intronic and retained introns). Three classes (ambiguous_orf, non-coding RNAs and TEC (to be experimentally confirmed) with small number of lncRNAs were merged into miscellaneous (misc) class for better representation. **(B)** Three major classes of differentially expressed lncRNAs identified in our study: intergenic (lincRNAs), processed transcripts and antisense lncRNAs. **(C)** The differences between the expression levels of top 3 most up-and down regulated protein-coding genes (in blue) and lncRNA genes (in red) in different the clinical were tested using Kruskal-Wallis with Dunn's multiple comparison test. The top 3 up-regulated protein-coding genes (*DUOXA2, CHI3L1* and *CXCL9*) and lncRNAs genes (*MMP12*, *FAM66D* and *SAA2-SAA4*), showed increasing signal intensity from control group to inflamed CD and UC groups based on averaged gene expression levels (p-value < 0.001). While, in case of the top 3 down-regulated protein-coding genes (*PCK1*, *GUCA2B* and *TNNC2*) and lncRNA genes (*DPP10-AS1, PDZK1P2* and *ANRIL*), a decreasing signal intensity across the clinical subgroups was observed from iCD, iUC to controls (p-value < 0.001). **(D)** A total of 8 genes were selected for real-time PCR validation of the microarray data in iCD vs control (red) and iUC vs control (blue). The log2 fold change (FC) is plotted on the y axis.

Additionally, we also tested the differences between the clinical subgroups for the top differentially expressed protein-coding and lncRNA genes. The top 3 up-regulated protein-coding genes (*DUOXA2, CHI3L1* and *CXCL9*) and lncRNAs genes (*MMP12*, *FAM66D* and *SAA2-SAA4*), showed increasing signal intensity based on the averaged gene expression levels across the spectrum of clinical subgroups from control to iCD and iUC (p-value < 0.001) (Figure 2C). In case of top 3 down-regulated protein-coding genes (*PCK1*, *GUCA2B* and *TNNC2*) and lncRNA genes (*DPP10-AS1, PDZK1P2* and *ANRIL*), we observed decreasing signal intensity across the clinical subgroups from iCD, iUC to controls (p-value < 0.001) (Figure 2C). Importantly, eight major isoforms (out of total 17 annotated isoforms) of *ANRIL* were found to be down-regulated in iCD and iUC when compared to controls and non-inflamed tissues in our data (Tables 4 and 5). *ANRIL* was −2.97 and −2.72 fold down-regulated in iCD vs control and iCD vs niCD contrasts, and −8.31 and −7.98 fold down-regulated in iUC vs control and iUC vs niUC contrasts, respectively.

Furthermore, for the validation of microarray results by quantitative real-time PCR (qPCR), we selected 8 top differentially expressed genes (based on FC) common between iCD vs control and iUC vs control (up-regulated: *DUOXA2, CHI3L1, DUOX2, MMP12, RP11-731 F5.2*; down-regulated: *PCK1, DPP10-AS1, ANRIL*). The qPCR analysis confirmed the microarray expression results with respect to the fold change values (Additional file 1: Table S5). We also performed qPCR analysis for *DUOX2* although it was not probed on our microarray, but it has been implicated along with its maturation factor *DUOXA2* in IBD pathogenesis (see discussion). Both *DUOXA2* and *DUOX2* were found to be significantly up-regulated in iCD vs control (FC = 8.83 and 5.85) and iUC vs control (FC = 9.14 and 6.05) (Figure 2D). In case of the remaining four contrasts, we also tested 5 differentially expressed genes by qPCR validation (Additional file 1: Table S6 and S7).

Overlap of differentially expressed genes in iCD and iUC

A Venn diagram illustrating the relationship between lncRNAs and protein-coding genes differentially expressed in iCD and iUC is shown in Figure 3. In total, 337 differentially expressed lncRNAs were identified as common between iCD and iUC with 100 unique lncRNAs for iCD and 400 unique lncRNAs for iUC (when compared to the healthy controls) (Figure 3A). While in case of the protein-coding genes, 901 differentially expressed genes were found to be common for iCD and iUC with 128 unique for iCD and 739 unique for iUC (Figure 3B). Conversely, in iCD vs iUC contrast, 19 out of 23 and 45 out of 50 differentially expressed lncRNAs and protein-coding genes, respectively overlapped with iCD vs control and iUC vs control.

Figure 3 Overlap of differentially expressed lncRNAs and protein-coding genes between iCD and iUC. Venn diagram shows an overlap of 337 lncRNAs **(A)** and 901 protein-coding genes **(B)** that were differentially expressed (fold change >1.5, adj.P-value <0.05) between patients with iCD and iUC when compared with healthy controls. We observed contra-regulated genes between iCD/iUC vs control contrasts when compared with iCD vs iUC contrast. The up-regulated genes are depicted in italics, down-regulated as underlined and contra-regulated in red. Heat maps of average normalized gene expression for the overlapping 337 lncRNAs **(C)** and 901 protein-coding genes **(D)** between iCD and iUC in the five clinical subgroups (iCD, niCD, iUC, niUC and controls) are displayed. Selected up-regulated and down-regulated genes are listed.

The unsupervised hierarchical clustering showed that both inflamed groups (iCD and iUC) cluster together in contrast to the non-inflamed (niCD and niUC) which clustered with healthy controls. The normalized gene expression values from the above mentioned 337 lncRNAs and 901 protein-coding genes common to both iCD and iUC conditions were averaged for each of the five clinical subgroups and visualized in a heat map in Figure 3C and D. The expression patterns for the specific up-regulated and down-regulated genes showed increasing or decreasing signal intensity across the clinical subgroups (from iCD, iUC, niCD, niUC and healthy controls). Collectively, these overlapping differentially expressed genes between iCD vs control and iUC vs control define a distinct inflammatory iCD/iUC gene expression signature. Importantly, this inflammatory gene signature included the key drivers of the innate and adaptive immune responses (for example *DUOXA2* and *CXCL1*).

Comparison of expression levels of top differentially expressed genes in patients and healthy controls

To stratify iCD and iUC samples from the healthy controls, we also compared the expression profiles of top 20 up/down-regulated lncRNAs and top 20 up/down-regulated protein-coding genes (based on FC) through unsupervised hierarchical clustering. The expression map of these top 40 differentially expressed genes displayed a clear separation of the patients from the control groups (Figure 4A and B), except for the two iUC samples B11 and 17_3 which were misclassified in the clustering. The magnitude of log2 intensity signal for these top differentially expressed genes displayed in Figure 4 was >6 in both iCD and iUC. Interestingly, in case of the iCD vs iUC contrast, clustering was unable to distinguish between iCD and iUC patients (Additional file 2: Figure S5). In addition to the top candidates, we also compared the expression profiles of all differentially expressed lncRNAs

and protein-coding genes, and observed similar results as described above (Additional file 2: Figure S6A and S6B).

Figure 4 Comparison of expression levels of the top 40 differentially expressed lncRNAs and protein-coding genes. Unsupervised hierarchical clustering of samples (patients in red, controls in blue) based on normalized expression values from the top 40 up and down-regulated lncRNAs and protein-coding genes for **(A)** iCD vs control and **(B)** iUC vs control. The log2 normalized expression values are shown in the color key. A clear separation between the diseased from control group is visible in case of iCD vs control contrast.

Inflammatory response and antimicrobial peptide (AMPs) genes are dysregulated in iCD and iUC

Antimicrobial peptides (AMPs) play an important role in protection of the host intestinal mucosa against microorganisms and AMPs dysregulation have been associated with IBD pathogenesis (see discussion for details). Therefore, we investigated whether there were differences in expression of genes involved in inflammatory response and AMPs, between different clinical subgroups. Our analysis identified key genes associated with inflammatory response, including the pro-inflammatory chemokines and cytokines. *CCL11*, *CCL19*, *CCL4* and *CXCL9*, were significantly up-regulated, in both iCD vs control and iUC vs control. In addition, we also found key antimicrobial response genes to be significantly up-regulated in iCD and iUC when compared to healthy controls (Figure 5). *REG3A*, *DEFA5* and *DEFA6* were >30 fold up-regulated only in iCD vs control. Chemokines *CXCL1* and *CXCL2* were >15 and >25 fold up-regulated in both iCD vs control and iUC vs control, respectively. *CXCL5*, *IL15* and *C3AR1* were specifically up-regulated in iCD vs control (Figure 5). Notably, *NOD2* gene was >2 fold up-regulated in iCD vs control, iUC vs control and iCD vs niCD contrasts. *DEFB1* and *NPY* were the only AMP genes that were significantly down-regulated in both iUC and iCD.

Figure 5 Differentially expressed protein-coding genes involved in antimicrobial and autoimmune response. Key genes involved autoimmune and inflammatory immune responses and AMPs were found to be dysregulated in both iCD and iUC when compared to healthy controls as well as to non-inflamed tissues. The log2 fold change values are plotted on the y axis.

Differentially expressed genes in iCD and iUC are enriched within IBD loci

Since majority of disease associated susceptibility SNPs map to the non-coding regions in the genome, we looked for the presence of known IBD associated SNPs (total 233 SNPs) within the Gencode v15 annotated lncRNAs. Interestingly, 29 IBD risk variants intersected 37 lncRNAs, of which only *IFNG-AS1* antisense lincRNA (*ENST00000536914.1*) harboring UC susceptibility SNP rs7134599 was found to be differentially expressed in our study. *IFNG-ASI* was up-regulated in iUC vs control (FC = 1.54) and iUC vs niUC (FC = 1.52). Furthermore, we identified IBD loci associated lncRNAs and protein-coding genes by intersecting the IBD susceptibility loci, which was defined as 500 kb long genomic region with the IBD risk variant in the middle. In total, 1040 IBD loci-associated lncRNAs were identified, out of which 96 lncRNAs were found to be differentially expressed (Additional file 1: Table S8). These differentially expressed lncRNAs co-localized with 57 IBD risk variants (within 500 kb locus), and were found to be enriched within IBD loci (p-value <

0.0001, Pearson's Chi-squared test). In case of protein-coding genes, 681 genes were found to be associated with IBD loci, out of which 154 were differentially expressed and enriched within IBD loci (p-value < 0.0001, Pearson's Chi-squared test). Based on unsupervised hierarchical clustering of averaged normalized gene expression values of 96 and 154 differentially expressed IBD loci-enriched lncRNAs and protein-coding genes, respectively, enabled independent stratification of disease from the controls and further distinguished inflamed from the non-inflamed conditions in both CD and UC (Figure 6).

Figure 6 Averaged gene expression for differentially expressed IBD loci-associated lncRNAs and protein-coding genes. Unsupervised hierarchical clustering of averaged normalized expression values for **(A)** 96 differentially expressed IBD loci-associated lncRNAs and **(B)** 154 protein-coding genes in different clinical subgroups. The range for the expression values is shown in the color scale.

Regulatory IBD-associated SNPs co-localize with differentially expressed IBD loci-associated lncRNAs

Next, we asked whether active regulatory regions within the IBD loci overlap with the differentially expressed IBD loci-associated lncRNAs. IBD associated SNPs overlapping active regulatory elements in intestinal epithelium were retrieved from Mokry et al. study [7]. In their study, the active regions overlapping IBD associated SNPs were identified based on H3K27ac chromatin immunoprecipitation and sequencing (ChIP-Seq). Out of 96 differentially expressed IBD loci-associated lncRNAs, 68 lncRNAs were found to be associated with 24 IBD loci SNPs co-localizing with the active regulatory elements in intestinal epithelium and immune cells (Additional file 1: Table S8). These overlapping IBD loci-associated active regulatory elements have been reported to frequently co-localize with the known transcription factor binding motifs [7]. A number of IBD-associated SNPs potentially affect the binding affinity of transcriptional factors, and thus perturb the gene expression. Additionally, IBD-associated risk variants also act as expression quantitative trait loci (eQTLs) signals for number of genes (Additional file 1: Table S8). For example, IBD-associated risk variant rs10797432 located within IBD loci-associated lncRNA *RP3-395 M20.8* (*ENSG00000238164)* alters the binding motifs for *TFAP2A* and *CTCF*. Furthermore, it is also known to acts as *cis*-eQTL for *MMEL1* in monocytes. Regulatory IBD-associated SNP rs1569723 located within IBD loci-associated lncRNA *RP11-465 L10.10* (*ENSG00000204044)* acts as *cis*-eQTL for *CD40* in monocytes. Also, SNP rs12946510 associated with lncRNAs *RP11-387H17.4* (*ENSG00000264968*) and *RP11-94 L15.2* (*ENSG00000264198*) is known to perturb the binding sites for transcription factors *FOXO1*, *ELF3*, and *SRF*. In addition, this SNP also acts as a *cis*-eQTL for pseudogene *KRT222P*, transcriptional co-activator complex component *MED24*, transcription factor *NR1D1*, and *ORMDL3* in lymphoblastoid cell lines. In case of the antisense lncRNA *CTD-2196E14.5 (ENSG00000261266)*, the associated SNP rs7404095 acts as a *cis*-eQTL for *PRKCB* in lymphoblastoid and *PRKCB1* in monocytes. Moreover, SNP rs734999 associated with lncRNA *RP3-395 M20.8 (ENSG00000238164)* acts as a *cis*-eQTL for *TNFRSF14* in lymphoblastoid cell line.

Cis-acting correlation of expression between differentially expressed IBD loci-associated lncRNAs and protein-coding genes

We computed pairwise Pearson correlations in order to explore the possible co-expression patterns between IBD loci-associated differentially expressed lncRNAs and protein-coding genes. Pairwise correlations of expression involving neighboring lncRNAs and protein-coding genes associated with each IBD-associated SNP (500 kb loci with SNP in the middle) were computed. We found positive ($r^2 \geq 0.5$) and extreme positive ($r^2 \geq 0.9$) correlations between the overlapping as well as *cis*-neighboring differentially expressed IBD loci-associated lncRNA-protein-coding gene pairs (p-value < 0.05). The pairwise correlations for six intersecting IBD loci associated lncRNA-protein-coding gene pairs *LSP1* and *ENST00000509204.1* (rs907611), *HLA-DQB1* and *ENST00000443574.1* (rs9268853, rs6927022), *MST1* and *ENST00000563780.1* (rs9822268 and rs3197999), *TSPAN33* and *ENST00000498745.1* (rs4728142), *SLC22A5* and *ENST00000417795.1* (rs2188962, rs12521868), *DGRD* and *ENST00000442524.1* (rs12994997, rs3792109) are plotted in Figure 7. Interestingly, lncRNA *ENST00000563780.1* and *MST1* protein-coding gene exhibited extreme positive correlation ($r^2 \geq 0.99$) (Figure 7). Enrichment for positive correlations has been reported for the lncRNAs intersecting protein-coding genes in antisense orientation [35]. Indeed, we also observed strong positive correlation ($r^2 \geq 0.7$) for the intersecting antisense lncRNA *ENST00000417795* and *SLC22A5* protein-coding gene.

Figure 7 Correlations of expression for *cis*-neighboring pairs of IBD loci-associated differentially expressed lncRNAs and protein-coding genes. Expression correlations for *cis*-neighboring pairs of IBD loci-associated lncRNAs and protein-coding genes. Overall positive correlations between overlapping protein-coding and lncRNA gene-pairs **(a)** *LSP1* and *ENST00000509204.1* **(b)** *HLA-DQB1* and *ENST00000443574.1* **(c)** *MST1* and *ENST00000563780.1* (d) *TSPAN33* and *ENST00000498745.1* **(e)** *SLC22A5* and *ENST00000417795.1* **(f)** *DGRD* and *ENST00000442524.1* were observed. An extreme positive ($r^2 \geq 0.99$, p-value < 2.2e-16) correlation was observed in case of *MST1* and its intersecting lncRNA *ENST00000563780.1* associated with IBD risk variants rs9822268 and rs3197999. Protein-coding expression is plotted on the x-axis, and lncRNA expression is shown on the y-axis. Each point represents a biopsy sample from different clinical subgroups.

Functional annotation of differentially expressed lncRNAs

The functional annotations of lncRNAs have mostly been based on nearest-neighbor approach i.e. "guilt-by-association" analyses, for example Cabili et al. [39]. We therefore analyzed the GO terms of genes that overlap with or are neighbors of the differentially expressed lncRNAs. We identified 516 nearest protein-coding neighbors within a span of <10 kb covering 610 differentially expressed lncRNAs. In addition, we also included 712 neighboring protein-coding genes for the 57 IBD risk variants (associated with 96 differentially expressed IBD loci-associated lncRNAs) based on 1 Mb locus size for each variant. The most significant over-represented GO terms in the biological process category included, antigen processing and presentation (p-value 7.39e-08), immune system process (p-value 2.5e-05) and natural killer cell activation (p-value 9.6e-05) (Additional file 1: Table S9). In the cellular component category, we found enrichment for MHC protein complex (p-value 5.95e-09). Furthermore, we also observed enrichment for over-represented GO terms in molecular function category which included protein binding, receptor binding and cytokine activity.

Cross validation of differentially expressed genes by SVM

Support Vector Machines (SVM) [33] was used for classifying IBD cases from controls and for cross-validating differentially expressed genes identified by LIMMA. The best SVM classifier performance was obtained from differentially expressed lncRNAs identified in iCD vs control followed by iUC vs control contrast (see methods for details). The classifier distinguished iCD and iUC from controls with 100% and 94.6% accuracy, 100% and 100% specificity and 100% and 86.7% sensitivity, respectively. In addition, the classifier was also able to distinguish iCD and iUC from niCD and niUC samples with an accuracy of 86.4% and 91.7%, with 78.3% and 88.9% specificity and 95.2% and 83.3% sensitivity, respectively. While for iCD vs iUC contrast, the accuracy of classifier was 77.8%, with 60.0% specificity and 90.4% sensitivity (Figure 8A). For the differentially expressed protein-coding genes, the classifier achieved accuracy of 100% and 94.6%, with 100% and 100% specificity, with %100 and 86.7% sensitivity, in discriminating iCD and iUC from controls, respectively (Figure 8B). Similar to the above described observations, the classifier also allowed distinction between iCD and iUC from niCD and niUC samples with an accuracy of 81.8% and 83.3%, with 78.2% and 77.8% specificity and 85.7% and 86.7% sensitivity, respectively. While for iCD vs iUC contrast, the accuracy of classifier was 88.9%, with 80.0% specificity and 95.2% sensitivity (Figure 8B). Furthermore, our classifier achieved the similar performance when using combined differentially expressed protein-coding and lncRNA genes or only protein-coding genes (data not shown). The effect of clinical parameters (Table 2 and Additional file 2: Figure S7) on disease (iCD or iUC) was described by the following linear function:

Figure 8 ROC curve analysis of differentially expressed lncRNAs and protein-coding genes. Receiver operating characteristic (ROC) curve analysis of differentially expressed **(A)** lncRNAs and **(B)** protein-coding genes for five contrasts.

$$y = 0.511 + (-0.212 * age) + (-0.114 * sex) + (-0.339 * smoking) + (1.185 * disease\ score) + (-0.058 * biopsy\ location)$$

Using t-statistics, p-values for linear regression coefficients for age, sex, smoking, disease index and biopsy location were 1.40e-01, 2.29e-01, 2.98e-03, 5.02e-07 and 6.43e-01, respectively. Our analysis indicated that disease index had strongest effect on defining iCD and iUC phenotype, followed by smoking, sex and age (Additional file 2: Figure S7, Figure S8). Differentially expressed genes identified by LIMMA were verified by Support Vector Machines- Recursive Feature Elimination (SVM-RFE) [35], which revealed a robust concordance rate in terms of total number of overlapping differentially expressed genes identified by the two methods (Additional file 2: Figure S9). In both iCD vs control, and iUC vs control contrasts, the overlap of about 66% was observed. To control for any input bias, a randomized lncRNA gene list of same size as differentially expressed lncRNAs was also used in this analysis (Additional file 2: Figure S8).

Impact of clinical parameters on disease diagnosis

We next investigated impact of the clinical parameters for disease diagnosis, and in the question if expression profiles of differentially expressed genes are also correlated to other clinical parameters. The applied strategies were linear regression model and weighted correlation network analysis (WGCNA) [36]. The regression analysis showed a strong impact of the disease index (Harvey-Bradshaw Index (HB) for CD and Simple Clinical Colitis

Activity Index (SCCAI) for UC) (p-value < 10e-6, t-test) which by definition is positively correlated to the severity of the disease, and a significant impact of smoking (p-value < 0.05, t-test) which was however, 3.5 times lower than the disease index and lower than the error rate. However, biopsy location did not show any significant effect on the severity of the disease. In agreement, the network analysis identified 10,435 genes significantly correlated to the disease index (p-value < 0.05, t-test), which is a clinical parameter with most related gene expression profiles. However, only 509 of these genes were differentially expressed between disease and control. Conversely, the expression profile of 1006 differentially expressed genes significantly associated with age (p-value < 0.05, t-test; Additional file 1: Table S10). These results suggest that even though the sample diagnosis for disease is only partly related to other clinical parameters, especially disease index and smoking, many differentially expressed genes in iCD and iUC also reflect impact of the patient's age. The average gene significance measures for all genes in a given module are summarized in Additional file 1: Table S10.

Overall, the network analysis identified three large co-expression modules enriched for differentially expressed genes between iCD / iUC and control (p-value < 10e-100, Pearson's Chi-squared test; Additional file 2: Figure S10 and S11). The three modules comprised of 2054 out of 2737 differentially expressed genes. The gene network of the "brown" module was found to be enriched for immune and pro-inflammatory response (Additional file 1: Table S11), the "green" and "red" module were driven by genes involved in small molecule trans-membrane transport, anionic and cationic transport (Additional file 1: Table S12 and S13). GO analysis was also performed for the randomized gene sets of the same module sizes. None of the randomized modules had significant GO terms.

Discussion

The present study was intended to explore the transcriptomic landscape of the lncRNAs in IBD, with particular focus on CD and UC. To explore the transcriptomic profiles of CD and UC patients, colonic pinch biopsies were analyzed using gene expression microarrays. Our results revealed widespread dysregulation of lncRNAs and protein-coding gene expression in both CD and UC. It is noteworthy that although our main focus was transcriptome analysis of lncRNAs, we also profiled a significant number of protein-coding genes (~12,000, see methods). The Gencode v15 lncRNA microarray has been extensively used and the levels of both mRNAs and lncRNAs are comparable and show strong correlations (ranging from 0.62 to 0.75) with results obtained from RNA sequencing (RNAseq) [38]. These correlations are also comparable with the previous lncRNA microarray versions [35]. The Gencode v15 lncRNA microarray has been designed to capture both poly(A) and non-poly(A) transcripts (out of total 22,007 lncRNA transcripts targeted by the microarray, 9273 lncRNA transcripts are polyadenylated). In recent years, many studies have been conducted to profile lncRNAs using RNAseq, however, it is expensive and time consuming because of the requirement of doing deep sequencing particularly lncRNAs which are relatively expressed at lower levels than the protein-coding genes [38]. It has also been reported that microarrays are more sensitive to detect whether a lncRNA is expressed or not as compared to RNAseq [40].

SVM based classifiers have been previously used to cross-validate the circulating microRNA based biomarker panels in UC [13]. We also verified robustness of the differentially expressed genes by SVM and the predictive capability of these genes to discriminate CD and UC was tested using SVM-RFE based classifiers. HB Index and SCCAI are symptom based indices used to assess the disease activity in CD and UC, respectively. Among various

clinical parameters tested, we found strong influence of disease index followed by smoking, age and sex, on iCD and iUC phenotypes. Smoking is known to have deleterious effects in CD while it has been found to be protective against UC [41]. Furthermore, smoking has also been known to influence colonic gene expression profile in CD [42]. Additionally, based on linear regression and WGCNA analysis, we did not find any significant effect of biopsy location on overall gene expression. However, regional variation in gene expression along the colonic mucosa has been reported to have influence on the expression profiling studies in IBD [43,44]. These modest regional variations are more pronounced in healthy controls and un-inflamed biopsies and largely remain masked when comparing inflamed biopsies [44]. On the contrary, other studies suggest no such gene expression differences due to regional variation [17,45]. These reports highlight the importance and impact of various confounding factors like smoking, sex, biopsy locations, among many other clinically relevant parameters in gene expression analysis in IBD.

Our analysis identified common expression patterns between the lncRNAs and protein-coding genes in iCD and iUC as confirmed by unsupervised hierarchical clustering (Figure 3). A distinctive inflammatory (iCD/iUC) gene expression signature included the key drivers of the innate and adaptive immune responses (chemokines, cytokines and defensins) for example *DUOXA2* [dual oxidase maturation factor 2], *CXCL1* [chemokine (C-X-C motif) ligand 1], *CXCL9* [chemokine (C-X-C motif) ligand 9] and also included a significant number of lncRNAs. Expression levels of both *DUOX2* and *DUOXA2* have been reported to be up-regulated in association with iUC, and in UC-CRC (UC-associated colorectal dysplasia and colorectal cancer) and are involved specifically in inflammation and regulated on a crypt-by-crypt basis in UC [46]. We also observed a global up-regulation of *DUOXA2* in iCD and iUC when compared to both non-inflamed and healthy controls. Both *DUOX2* and its maturation factor *DUOXA2* are part of the *NADPH* oxidase family of enzymes involved in release of hydrogen peroxide (H_2O_2) [47]. These enzymes are essential components of evolutionarily conserved mechanisms through which organisms are known to defend themselves against bacterial, viral, or parasitic infections, yet allowing tolerance of commensals [48,49] Suppression of *DUOX2*-generated H_2O_2 production by Mesalazine (5-aminosalicylic acid; 5-ASA) has been demonstrated to reduce reactive oxygen species (ROS)-induced genetic lesions and thereby lowering the risk of UC-CRC [46].

Our results revealed significant down-regulation of lncRNA *ANRIL* [antisense non-coding RNA in the INK4 locus] in both iCD (FC < −2.7, p-value < 0.05) and iUC (FC < −7.9, p-value < 0.05) when compared to non-inflamed and healthy controls. *ANRIL*, encoded on the chromosome 9p2.3 region is a known hotspot for the disease-associated SNPs [50]. *ANRIL* has emerged as an important regulatory molecule mediating human disease at various levels and cellular settings. Nevertheless, the role of *ANRIL* has not yet been described specifically in context of the IBD pathology. *ANRIL* has been found to be up-regulated in leukemia, prostate cancer, basal cell carcinoma and glioma, whereas, depletion of *ANRIL* has been implicated with reduced proliferation, indicating its role in cancerogenesis [51-53]. Remarkably, in our study, eight major *ANRIL* isoforms, including the isoforms known to form circular variants (*cANRIL*), were found to be universally down-regulated in both iCD and iUC. Importantly, endogenous expression of *cANRIL* has been associated with risk for atherosclerosis [54]. In this context, dysregulation of *ANRIL* in IBD is highly intriguing, particularly the down-regulation of *cANRIL* isoform. Indeed, recently, circular RNAs have been shown to be involved in stabilizing the sense transcripts and also act as sponges for microRNAs [55] however; biological functions of circRNAs have recently been debated [56].

It is therefore imperative to investigate comprehensively the potential roles of *cANRIL* in IBD pathogenesis.

Unsurprisingly, our results also enabled us to distinguish between iCD and iUC, although the number of differentially expressed genes were small, which emphasizes the close pathogenic nature of CD and UC. An interesting distinction between iCD and iUC expression was observed in *SERPINB3* [serpin peptidase inhibitor, clade B (ovalbumin), member 3] expression, where *SERPINB3* was significantly down-regulated (FC < −3.8) in iCD vs iUC contrast (Table 6 and Figure 1G). *SERPINB3* has been found to be over-expressed in certain squamous epithelial cancers, for instance uterine cervix carcinoma, head and neck carcinomas, and esophagus carcinoma [57] Although, the precise physiological functions of *SERPINB3* are elusive, it has been hypothesized that *SERPINB3* might be involved in the development of autoimmunity [58].

In our study, we found significant enrichment for the 96 differentially expressed lncRNAs within IBD loci. Collectively, we found differentially expressed IBD loci-associated lncRNAs overlapping active regulatory elements in intestinal epithelium and immune cells located within known binding motifs [7]. LncRNA *RP3-395 M20.8* was found to be associated with the regulatory IBD risk variant rs10797432 which affects the binding motifs for *AP-2* [transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)] and *CTCF* [CCCTC-binding factor (zinc finger protein)] (Additional file 1: Table S5). Moreover, IBD risk variant rs1569723 is known to act as a *cis*-eQTL for *CD40* [CD40 molecule, TNF receptor superfamily member 5] which was significantly up-regulated in iUC and associated with lncRNA *RP11-465 L10.10.* Additionally, lncRNA *IFNG-ASI* harboring UC susceptibility SNP rs7134599 was found to be up-regulated in iUC. SNP rs7134599 is associated with IBD26 (12q15) genetic locus and with regulatory pro-inflammatory cytokines *IFNG* [interferon, gamma] and *IL-2* [interleukin 2] and anti-inflammatory cytokine *IL-26* [interleukin 26]. *IFNG* gene encodes interferon gamma (*IFN*-γ), a soluble cytokine that is pivotal for the host's innate and adaptive immunity against viral, certain bacterial and protozoal infections. Aberrant expression of *IFN*-γ has been linked with a number of autoimmune and inflammatory diseases, and mucosal expression of *IFN*-γ is known to play a vital role in pathogenesis IBD [59]. IL-2 is encoded by *IL2* gene and is involved in immune responses to microbial infections and intestinal inflammation activation in IBD. Anti-inflammatory *IL-26* has been shown to be overexpressed in CD [60]. These findings suggest potential involvement of differentially expressed lncRNAs overlapping the active regulatory elements in IBD pathogenesis.

Interestingly, we also found positive ($r^2 \geq 0.5$) and extreme positive ($r^2 \geq 0.9$) correlations between the overlapping as well as *cis*-neighboring differentially expressed IBD loci-associated lncRNA-protein-coding gene pairs. A strong positive correlation was observed between lncRNA *AC051649.12* and protein-coding gene *LSP1* [lymphocyte-specific protein 1] associated with IBD risk variant rs907611. SNP rs907611 affects the binding affinity of transcriptional factors *YY1* and *NF-muE1* and thus alters the gene expression. It is plausible that the differentially expressed IBD loci-associated lncRNAs intersecting protein-coding genes somehow contribute to the regulation of the latter [61]. Taken together, these data suggests role of lncRNAs in regulating the expression of IBD loci-associated genes.

Additionally, we also noticed dysregulation of AMPs and inflammatory response genes such as, pro-inflammatory chemokines and cytokines in various clinical subgroups. For example, key antimicrobial response genes, *REG3A* (Regenerating islet-derived 3 alpha), *DEFA5*

(Defensin, alpha 5, Paneth cell-specific) and *DEFA6* (Defensin, alpha 6, Paneth cell-specific), were >30 fold up-regulated specially in iCD vs control. Consistent with our results, *REG3A*, *DEFA5* and *DEFA6*, have been shown previously to be significantly up-regulated and linked to Paneth cell metaplasia in IBD [62,63]. Mutations in the cytoplasmic pathogen recognition receptor *NOD2* (nucleotide-binding oligomerization domain containing 2) gene have been associated with ileal CD and Paneth cell dysfunction [64] and importantly, *NOD2* was found to be up-regulated in both iCD and iUC. Concordant with the findings by Arjis et al., we also found two AMPs *DEFB1 and NPY* significantly down-regulated in both iCD and iUC [59]. *IL15* (interleukin 15) was found specifically up-regulated in iCD and not in iUC which supports the notion that it contributes to acute intestinal inflammation in CD [65].

For all the differentially expressed lncRNAs and protein-coding genes, we evaluated biological functional processes through analysis of GO terms based on "guilt-by-association" and WGCNA approach. Unsurprisingly, we found enrichment for immune response, pro-inflammatory cytokine activity, extracellular matrix organization, and ion membrane transport genes (Additional file 1: Table S9, S11, S12 and S13). Given the idiopathic nature of IBD, the overall up-regulation of pro-inflammatory immune response-related gene expression manifestation could be largely due to the infiltrating immune cells, rather than due to the underlying disease phenotype. Indeed, persistent inflammation in CD and UC is known to be elicited by the activation of innate and adaptive immune cells by foreign antigens, which in turn produce and release pro-inflammatory cytokines that give rise to the acrimonious circle of inflammation thereby leading to chronic tissue injury and epithelial damage [66]. Nevertheless, differentially expressed genes identified in non-inflamed (niCD and niUC) vs control (Additional file 1: Table S3 and S4), might be disease specific. In summary, our findings suggest that dysregulated lncRNAs could be involved in the IBD pathogenesis. However, these findings warrant a systematic experimental follow-up in cellular and murine based models with additional validation in a larger cohort in order to elucidate the role and biomarker potential of these dysregulated lncRNAs in IBD.

Conclusions

In conclusion, we show that lncRNA expression profiling can be effectively used to stratify iCD and iUC from healthy controls. Additionally, our data indicates the underlining potential of lncRNA transcriptional signatures associated with clinical parameters as biomarkers for IBD.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Conceived and designed the experiments: FP, JG, CHBB, AHM, MV, KSF. Performed the experiments: AHM, CHBB, SES, XP. Analyzed the data and wrote the manuscript: AHM. Intellectual guidance: FP, JG. All authors read and approved the final manuscript.

Acknowledgements

We thank the Genomics Unit's microarray core facility at Centre for Genomic Regulation (CRG), Barcelona, Spain, for performing the microarray experiments. We thank Sarah Bonin and Rory Johnson for helpful advice for initial QC of the data and Marc R. Friedlander for help with sample logistics at CRG. We also would like to thank Lene Buhl Riis for histological evaluation of biopsies and Lene Normann Nielsen and Helle Buck Rasmussen for sample collection and handling.

Funding

This work was supported by grants from Danish Council for Strategic Research, Danish Council for Independent Research (Technology and Production Sciences), Danish Center for Scientific Computing (DCSC, DeiC) and Poul and Erna Sehested Hansen Foundation for the funding. MV is supported by Dagny Sigismunde Marie Sofie Rasmine Hertz Johansens and Andreas Hermann Johansens Fond. XP is supported by Innovation Fund Denmark. The funders had no role in study design, data collection and analysis or preparation of the manuscript.

References

1. Van Limbergen J, Radford-Smith G, Satsangi J. Advances in IBD genetics. Nat Rev Gastroenterol Hepatol. 2014;11:372–85.

2. Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. Nature. 2007;448:427–34.

3. Abraham C, Cho JH. Inflammatory bowel disease. N Engl J Med. 2009;361:2066–78.

4. Sartor RB. Genetics and environmental interactions shape the intestinal microbiome to promote inflammatory bowel disease versus mucosal homeostasis. Gastroenterology. 2010;139:1816–9.

5. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491:119–24.

6. Mirza AH, Kaur S, Brorsson CA, Pociot F. Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. PLoS One. 2014;9:e105723.

7. Mokry M, Middendorp S, Wiegerinck CL, Witte M, Teunissen H, Meddens CA, et al. Many inflammatory bowel disease risk loci include regions that regulate gene expression in immune cells and the intestinal epithelium. Gastroenterology. 2014;146:1040–7.

8. Pekow JR, Kwon JH. MicroRNAs in inflammatory bowel disease. Inflamm Bowel Dis. 2012;18:187–93.

9. Granlund A, Flatberg A, Østvik AE, Drozdov I, Gustafsson BI, Kidd M, et al. Whole genome gene expression meta-analysis of inflammatory bowel disease colon mucosa

demonstrates lack of major differences between Crohn's disease and ulcerative colitis. PLoS One. 2013;8:e56818.

10. Coskun M, Bjerrum JT, Seidelin JB, Nielsen OH. MicroRNAs in inflammatory bowel disease–pathogenesis, diagnostics and therapeutics. World J Gastroenterol. 2012;18:4629–34.

11. Lin J, Welker NC, Zhao Z, Li Y, Zhang J, Reuss SA, et al. Novel specific microRNA biomarkers in idiopathic inflammatory bowel disease unrelated to disease activity. Mod Pathol. 2014;27:602–8.

12. Iborra M, Bernuzzi F, Invernizzi P, Danese S. MicroRNAs in autoimmunity and inflammatory bowel disease: crucial regulators in immune response. Autoimmun Rev. 2012;11:305–14.

13. Duttagupta R, DiRienzo S, Jiang R, Bowers J, Gollub J, Kao J, et al. Genome-wide maps of circulating miRNA biomarkers for ulcerative colitis. PLoS One. 2012;7:e31241.

14. Haberman Y, Tickle TL, Dexheimer PJ, Kim M-O, Tang D, Karns R, et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. J Clin Invest. 2014;124:3617–33.

15. Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, et al. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. Nat Genet. 2011;43:242–5.

16. McKenna LB, Schug J, Vourekas A, McKenna JB, Bramswig NC, Friedman JR, et al. MicroRNAs control intestinal epithelial differentiation, architecture, and barrier function. Gastroenterology. 2010;139:1654–64.

17. Wu F, Dassopoulos T, Cope L, Maitra A, Brant SR, Harris ML, et al. Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: insights into distinctive pathogenesis. Inflamm Bowel Dis. 2007;13:807–21.

18. Olsen J, Gerds TA, Seidelin JB, Csillag C, Bjerrum JT, Troelsen JT, et al. Diagnosis of ulcerative colitis before onset of inflammation by multivariate modeling of genome-wide gene expression data. Inflamm Bowel Dis. 2009;15:1032–8.

19. Montero-Meléndez T, Llor X, García-Planella E, Perretti M, Suárez A. Identification of novel predictor classifiers for inflammatory bowel disease by gene expression profiling. PLoS One. 2013;8:e76235.

20. Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge R-M, Chang HY. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. Elife. 2013;2:e00762.

21. Geboes K, Riddell R, Ost A, Jensfelt B, Persson T, Löfberg R. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. Gut. 2000;47:404–9.

22. D'Haens GR, Geboes K, Peeters M, Baert F, Penninckx F, Rutgeerts P. Early lesions of recurrent Crohn's disease caused by infusion of intestinal contents in excluded ileum. Gastroenterology. 1998;114:262–7.

23. GENCODE Custom lncRNA Expression Microarray Design. www.gencodegenes.org/lncrna_microarray.html. Accessed 2 August 2014.

24. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, et al. A comparison of background correction methods for two-colour microarrays. Bioinformatics. 2007;23:2700–7.

25. Bolstad B. Probe Level Quantile Normalization of High Density Oligonucleotide Array Data. http://bmbolstad.com/stuff/qnorm.pdf. Accessed 20 October 2014.

26. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3:Article3.

27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Statistical Soc Series B. 1995;57:289–300.

28. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5:R80. http://cran.r-project.org/.

29. ImmunoBase. http://www.immunobase.org. Accessed 30 October 2014.

30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

31. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22:1790–7.

32. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. 2013;41:D377–86. http://www.pantherdb.org/.

33. Vapnik VN. The nature of statistical learning theory. 2nd edition. Statistics for engineering and information science. New York: Springer; 2000.

34. Pedregosa F, Varoquaux G, Gramfort A, Michel A, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Machine Learning Res. 2011;12:2825–30.

35. Duan K-B, Rajapakse JC, Wang H, Azuaje F. Multiple SVM-RFE for gene selection in cancer classification with expression data. IEEE Trans Nanobioscience. 2005;4:228–34.

36. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

37. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics. 2007;23:257–8.

38. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22:1775–89.

39. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011;25:1915–27.

40. Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC Genomics. 2009;10:161.

41. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011;474:307–17.

42. Nielsen OH, Bjerrum JT, Csillag C, Nielsen FC, Olsen J. Influence of smoking on colonic gene expression profile in Crohn's disease. PLoS One. 2009;4:e6210.

43. LaPointe LC, Dunne R, Brown GS, Worthley DL, Molloy PL, Wattchow D, et al. Map of differential transcript expression in the normal human large intestine. Physiol Genomics. 2008;33:50–64.

44. Noble CL, Abbas AR, Cornelius J, Lees CW, Ho G-T, Toy K, et al. Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. Gut. 2008;57:1398–405.

45. Costello CM, Mah N, Häsler R, Rosenstiel P, Waetzig GH, Hahn A, et al. Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. PLoS Med. 2005;2:e199.

46. MacFie TS, Poulsom R, Parker A, Warnes G, Boitsova T, Nijhuis A, et al. DUOX2 and DUOXA2 form the predominant enzyme system capable of producing the reactive oxygen species H2O2 in active ulcerative colitis and are modulated by 5-aminosalicylic acid. Inflamm Bowel Dis. 2014;20:514–24.

47. Bedard K, Krause K-H. The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. Physiol Rev. 2007;87:245–313.

48. Bae YS, Choi MK, Lee W-J. Dual oxidase in mucosal immunity and host-microbe homeostasis. Trends Immunol. 2010;31:278–87.

49. Ha E-M, Lee K-A, Seo YY, Kim S-H, Lim J-H, Oh B-H, et al. Coordination of multiple dual oxidase-regulatory pathways in responses to commensal and infectious microbes in drosophila gut. Nat Immunol. 2009;10:949–57.

50. Pasmant E, Sabbagh A, Vidaud M, Bièche I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. FASEB J. 2011;25:444–8.

51. Sherborne AL, Hosking FJ, Prasad RB, Kumar R, Koehler R, Vijayakrishnan J, et al. Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. Nat Genet. 2010;42:492–4.

52. Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell. 2010;38:662–74.

53. Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. PLoS Genet. 2010;6:e1000899.

54. Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. PLoS Genet. 2010;6:e1001233.

55. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. Nature. 2013;495:384–8.

56. Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. Genome Biol. 2014;15:409.

57. Izuhara K, Ohta S, Kanaji S, Shiraishi H, Arima K. Recent progress in understanding the diversity of the human ov-serpin/clade B serpin family. Cell Mol Life Sci. 2008;65:2541–53.

58. Vidalino L, Doria A, Quarta S, Zen M, Gatta A, Pontisso P. SERPINB3, apoptosis and autoimmunity. Autoimmun Rev. 2009;9:108–12.

59. Neurath MF. Cytokines in inflammatory bowel disease. Nat Rev Immunol. 2014;14:329–42.

60. Dambacher J, Beigel F, Zitzmann K, De Toni EN, Göke B, Diepolder HM, et al. The role of the novel Th17 cytokine IL-26 in intestinal inflammation. Gut. 2009;58:1207–17.

61. Gingeras TR. Origin of phenotypes: genes and transcripts. Genome Res. 2007;17:682–90.

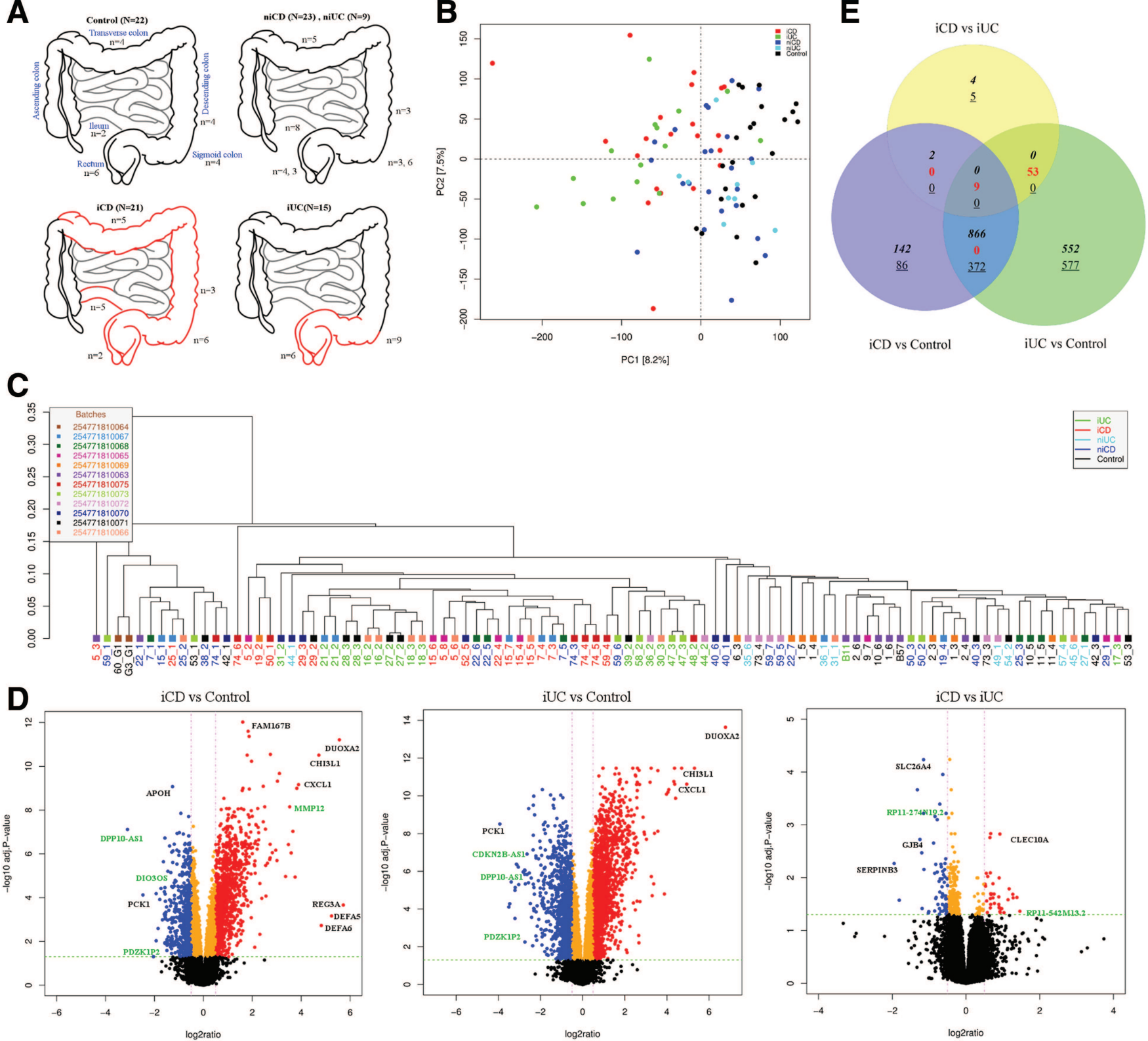
62. Arijs I, De Hertogh G, Lemaire K, Quintens R, Van Lommel L, Van Steen K, et al. Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment. PLoS One. 2009;4:e7984.

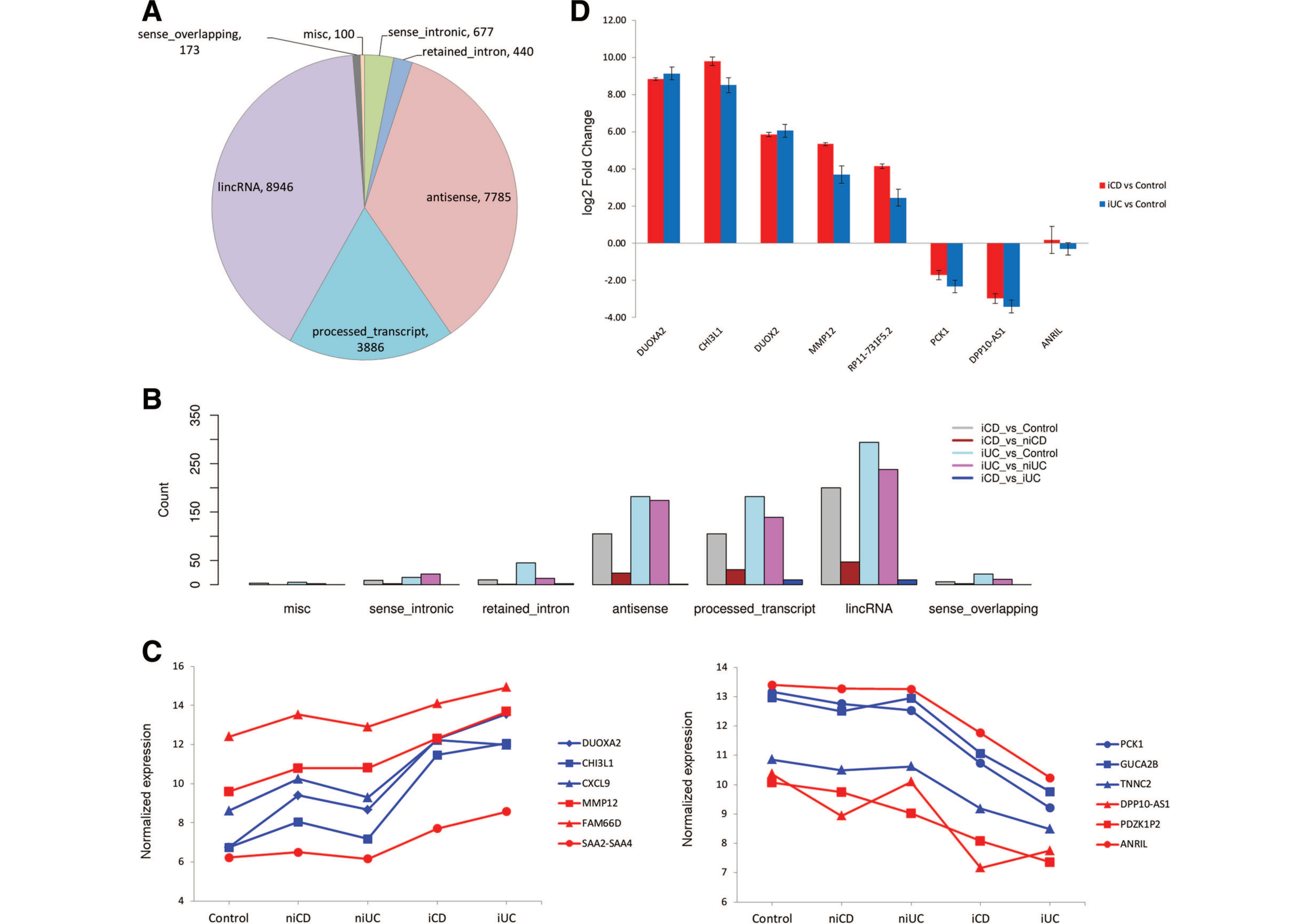
63. Wehkamp J, Schmid M, Stange EF. Defensins and other antimicrobial peptides in inflammatory bowel disease. Curr Opin Gastroenterol. 2007;23:370–8.

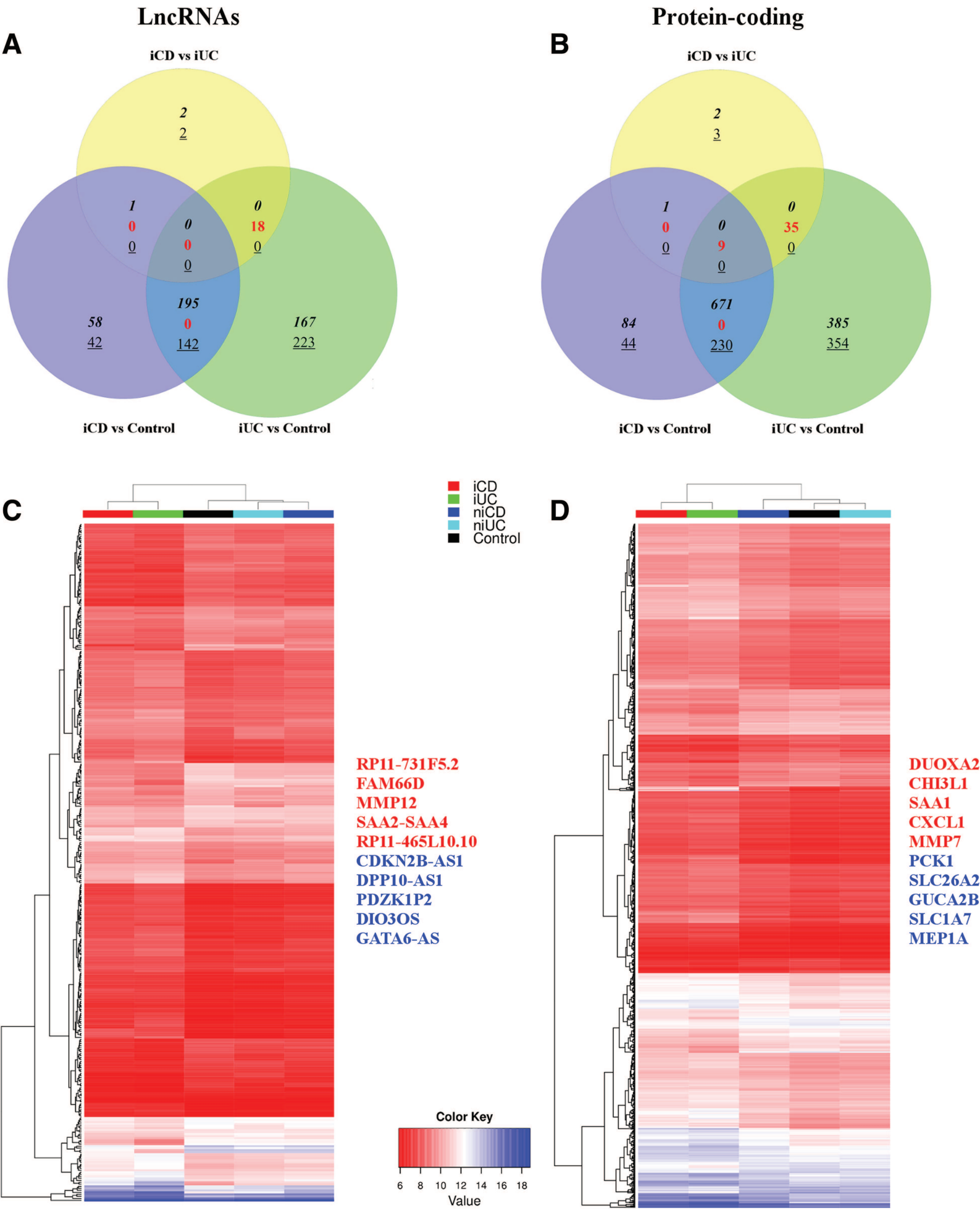
64. Hamm CM, Reimers MA, McCullough CK, Gorbe EB, Lu J, Gu CC, et al. NOD2 status and human ileal gene expression. Inflamm Bowel Dis. 2010;16:1649–57.

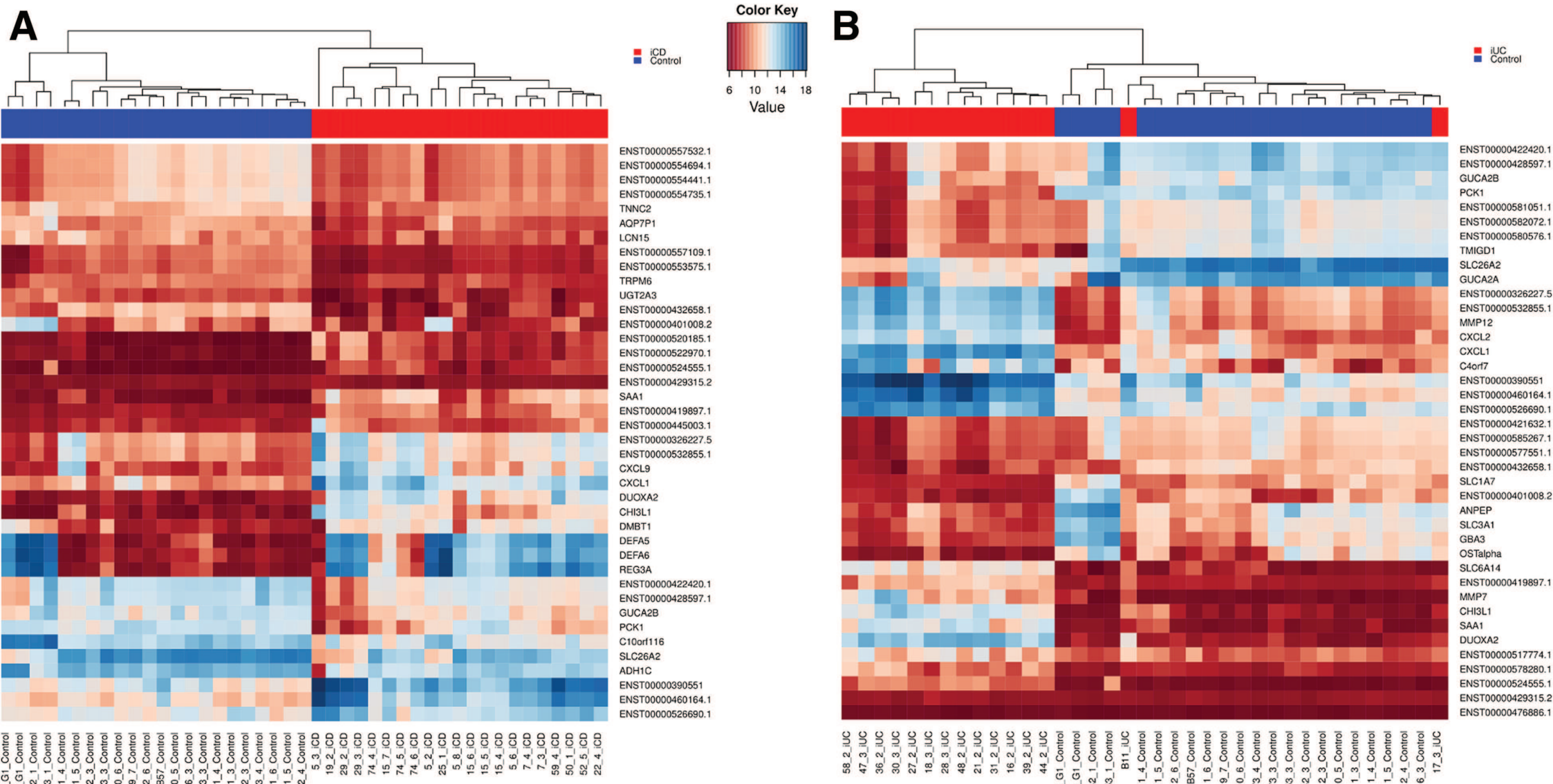
65. Schulthess J, Meresse B, Ramiro-Puig E, Montcuquet N, Darche S, Bègue B, et al. Interleukin-15-dependent NKp46+ innate lymphoid cells control intestinal inflammation by recruiting inflammatory monocytes. Immunity. 2012;37:108–21.

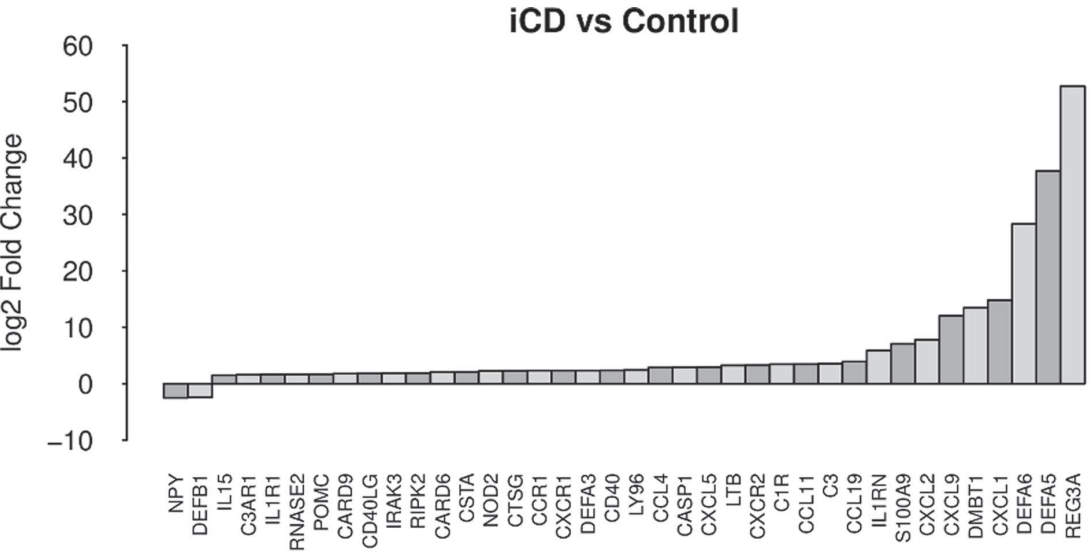
66. Bouma G, Strober W. The immunological and genetic basis of inflammatory bowel disease. Nat Rev Immunol. 2003;3:521–33.

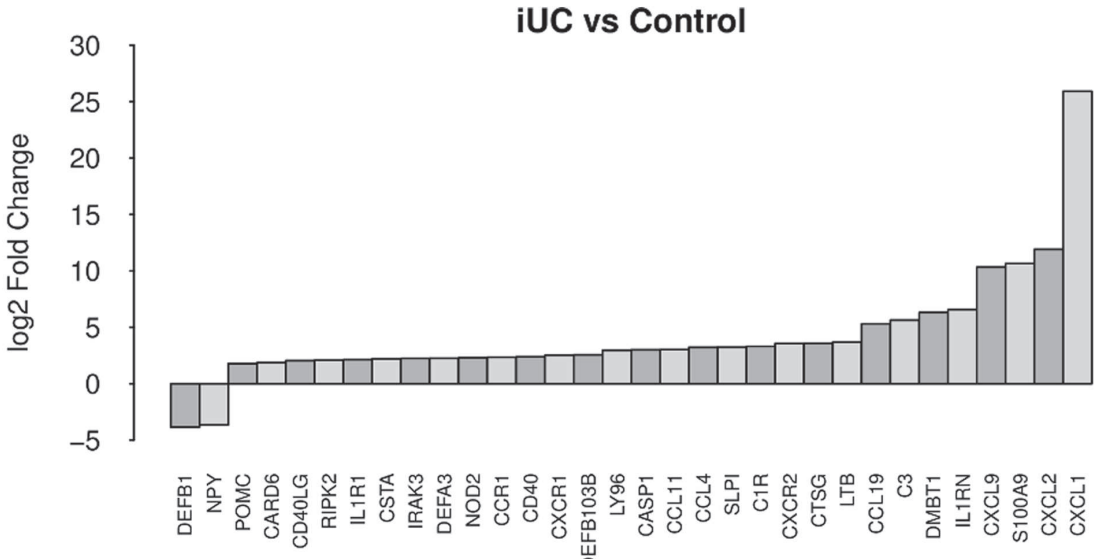

A
Control (N=22)
Transverse colon
n=4
Ascending colon
Descending colon
n=4
Ileum
n=2
Rectum
n=6
Sigmoid colon
n=4
niCD (N=23) , niUC (N=9)
n=5
n=8
n=3
n=4, 3
n=3, 6
iCD (N=21)
n=5
n=3
n=5
n=6
n=2
iUC(N=15)
n=6
n=9
B
PC2 [7.5%]
PC1 [8.2%]
iCD
iUC
niCD
niUC
Control
E
iCD vs iUC
4
5
2
0
0
0
9
0
0
53
0
866
0
372
142
86
552
577
iCD vs Control
iUC vs Control
C
Batches
254771810064
254771810067
254771810068
254771810065
254771810069
254771810063
254771810075
254771810073
254771810072
254771810070
254771810071
254771810066
iUC
iCD
niUC
niCD
Control
D
iCD vs Control
FAM167B
DUOXA2
CHI3L1
CXCL1
APOH
MMP12
DPP10-AS1
DIO3OS
PCK1
REG3A
DEFA5
DEFA6
PDZK1P2
−log10 adj.P-value
log2ratio
iUC vs Control
DUOXA2
CHI3L1
CXCL1
PCK1
CDKN2B-AS1
DPP10-AS1
PDZK1P2
−log10 adj.P-value
log2ratio
iCD vs iUC
SLC26A4
RP11-274N19.2
GJB4
CLEC10A
SERPINB3
RP11-542M13.2
−log10 adj.P-value
log2ratio

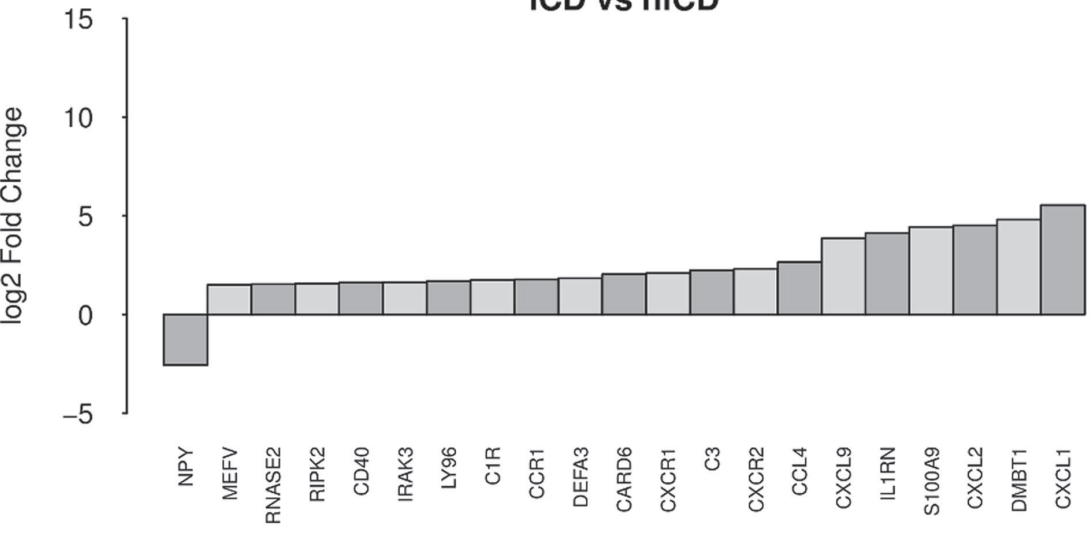

A
sense_overlapping, 173
misc, 100
sense_intronic, 677
retained_intron, 440
lincRNA, 8946
antisense, 7785
processed_transcript, 3886
D
log2 Fold Change
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-2.00
-4.00
DUOXA2
CHI3L1
DUOX2
MMP12
RP11-731F5.2
PCK1
DPP10-AS1
ANRIL
iCD vs Control
iUC vs Control
B
Count
0
50
150
250
350
misc
sense_intronic
retained_intron
antisense
processed_transcript
lincRNA
sense_overlapping
iCD_vs_Control
iCD_vs_niCD
iUC_vs_Control
iUC_vs_niUC
iCD_vs_iUC
C
Normalized expression
16
14
12
10
8
6
4
Control
niCD
niUC
iCD
iUC
DUOXA2
CHI3L1
CXCL9
MMP12
FAM66D
SAA2-SAA4
Normalized expression
14
13
12
11
10
9
8
7
6
Control
niCD
niUC
iCD
iUC
PCK1
GUCA2B
TNNC2
DPP10-AS1
PDZK1P2
ANRIL

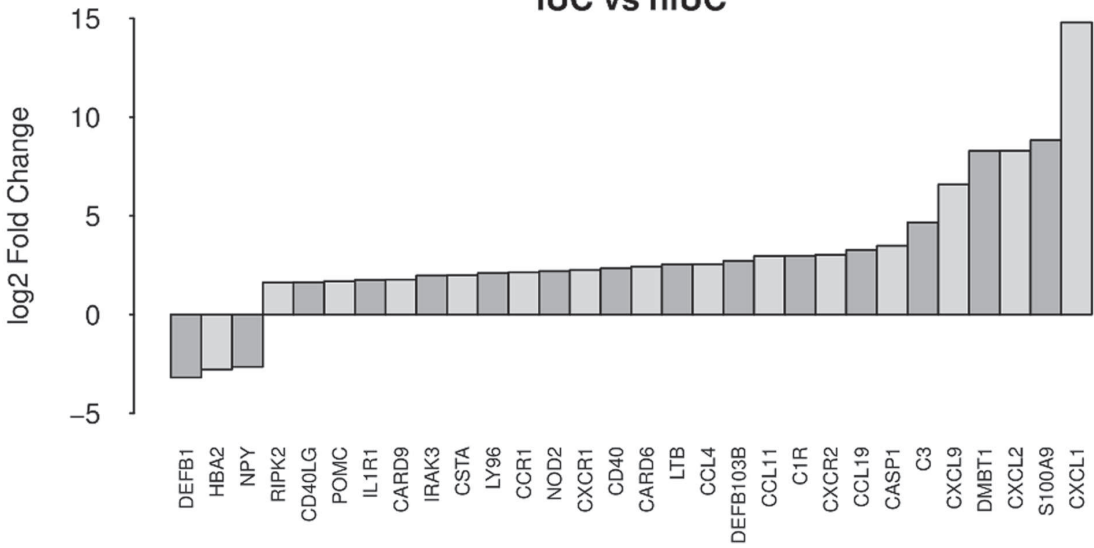

LncRNAs
A
iCD vs iUC
2
2
1
0
0
0
0
0
0
18
0
195
0
142
58
42
167
223
iCD vs Control
iUC vs Control
Protein-coding
B
iCD vs iUC
2
3
1
0
0
0
0
0
0
35
0
671
0
230
84
44
385
354
iCD vs Control
iUC vs Control
C
iCD
iUC
niCD
niUC
Control
RP11-731F5.2
FAM66D
MMP12
SAA2-SAA4
RP11-465L10.10
CDKN2B-AS1
DPP10-AS1
PDZK1P2
DIO3OS
GATA6-AS
Color Key
6 8 10 12 14 16 18
Value
D
DUOXA2
CHI3L1
SAA1
CXCL1
MMP7
PCK1
SLC26A2
GUCA2B
SLC1A7
MEP1A

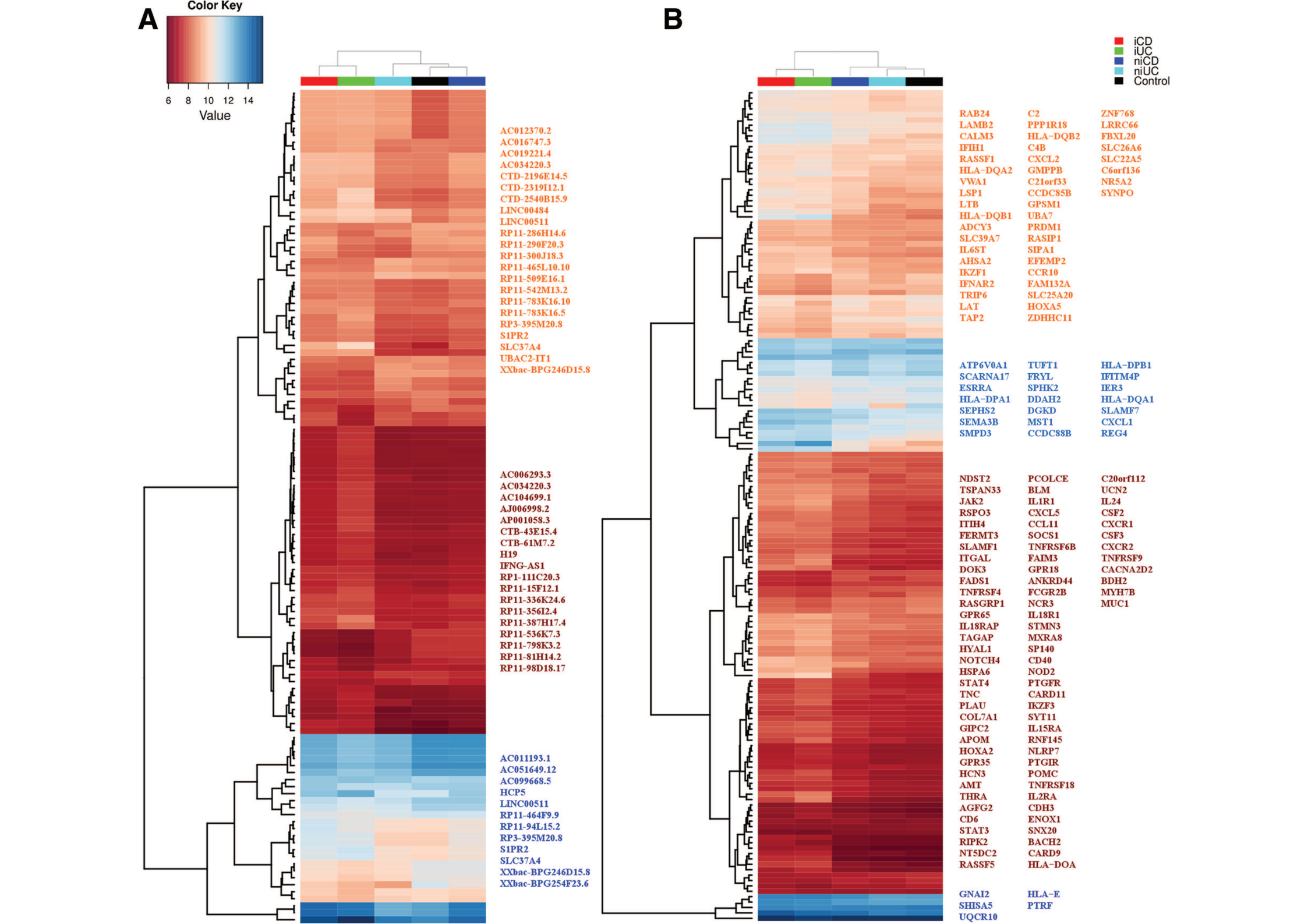

A
iCD
Control
Color Key
6
10
14
18
Value
ENST00000557532.1
ENST00000554694.1
ENST00000554441.1
ENST00000554735.1
TNNC2
AQP7P1
LCN15
ENST00000557109.1
ENST00000553575.1
TRPM6
UGT2A3
ENST00000432658.1
ENST00000401008.2
ENST00000520185.1
ENST00000522970.1
ENST00000524555.1
ENST00000429315.2
SAA1
ENST00000419897.1
ENST00000445003.1
ENST00000326227.5
ENST00000532855.1
CXCL9
CXCL1
DUOXA2
CHI3L1
DMBT1
DEFA5
DEFA6
REG3A
ENST00000422420.1
ENST00000428597.1
GUCA2B
PCK1
C10orf116
SLC26A2
ADH1C
ENST00000390551
ENST00000460164.1
ENST00000526690.1
60_G1_Control
G3_G1_Control
42_1_Control
53_1_Control
11_4_Control
11_5_Control
42_3_Control
73_3_Control
10_6_Control
9_7_Control
2_6_Control
B57_Control
10_5_Control
6_3_Control
53_3_Control
1_4_Control
1_3_Control
2_3_Control
73_4_Control
1_6_Control
1_5_Control
2_4_Control
5_3_iCD
19_2_iCD
29_2_iCD
29_3_iCD
74_4_iCD
15_7_iCD
74_5_iCD
74_6_iCD
5_2_iCD
25_1_iCD
5_8_iCD
15_6_iCD
15_5_iCD
15_4_iCD
5_6_iCD
7_4_iCD
7_3_iCD
59_4_iCD
50_1_iCD
52_5_iCD
22_4_iCD
B
iUC
Control
ENST00000422420.1
ENST00000428597.1
GUCA2B
PCK1
ENST00000581051.1
ENST00000582072.1
ENST00000580576.1
TMIGD1
SLC26A2
GUCA2A
ENST00000326227.5
ENST00000532855.1
MMP12
CXCL2
CXCL1
C4orf7
ENST00000390551
ENST00000460164.1
ENST00000526690.1
ENST00000421632.1
ENST00000585267.1
ENST00000577551.1
ENST00000432658.1
SLC1A7
ENST00000401008.2
ANPEP
SLC3A1
GBA3
OSTalpha
SLC6A14
ENST00000419897.1
MMP7
CHI3L1
SAA1
DUOXA2
ENST00000517774.1
ENST00000578280.1
ENST00000524555.1
ENST00000429315.2
ENST00000476886.1
58_2_iUC
47_3_iUC
36_2_iUC
30_3_iUC
27_2_iUC
18_3_iUC
28_3_iUC
48_2_iUC
21_2_iUC
31_2_iUC
16_2_iUC
39_2_iUC
44_2_iUC
G3_G1_Control
60_G1_Control
42_1_Control
53_1_Control
B11_iUC
11_4_Control
11_5_Control
2_6_Control
B57_Control
1_6_Control
9_7_Control
10_6_Control
73_4_Control
73_3_Control
53_3_Control
42_3_Control
2_3_Control
10_5_Control
1_3_Control
1_4_Control
1_5_Control
2_4_Control
6_3_Control
17_3_iUC

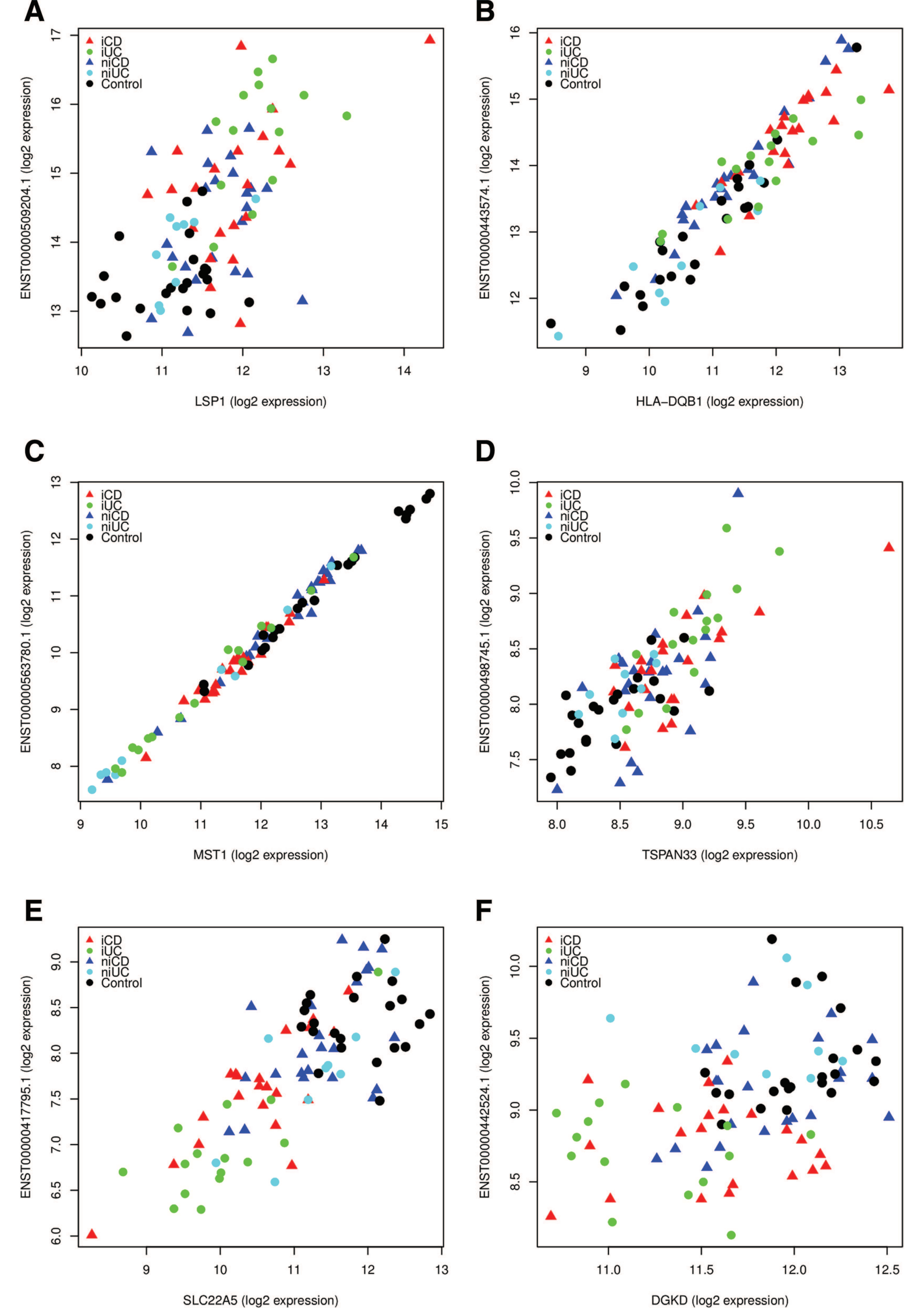

iCD vs Control
log2 Fold Change
60
50
40
30
20
10
0
−10
NPY
DEFB1
IL15
C3AR1
IL1R1
RNASE2
POMC
CARD9
CD40LG
IRAK3
RIPK2
CARD6
CSTA
NOD2
CTSG
CCR1
CXCR1
DEFA3
CD40
LY96
CCL4
CASP1
CXCL5
LTB
CXCR2
C1R
CCL11
C3
CCL19
IL1RN
S100A9
CXCL2
CXCL9
DMBT1
CXCL1
DEFA6
DEFA5
REG3A

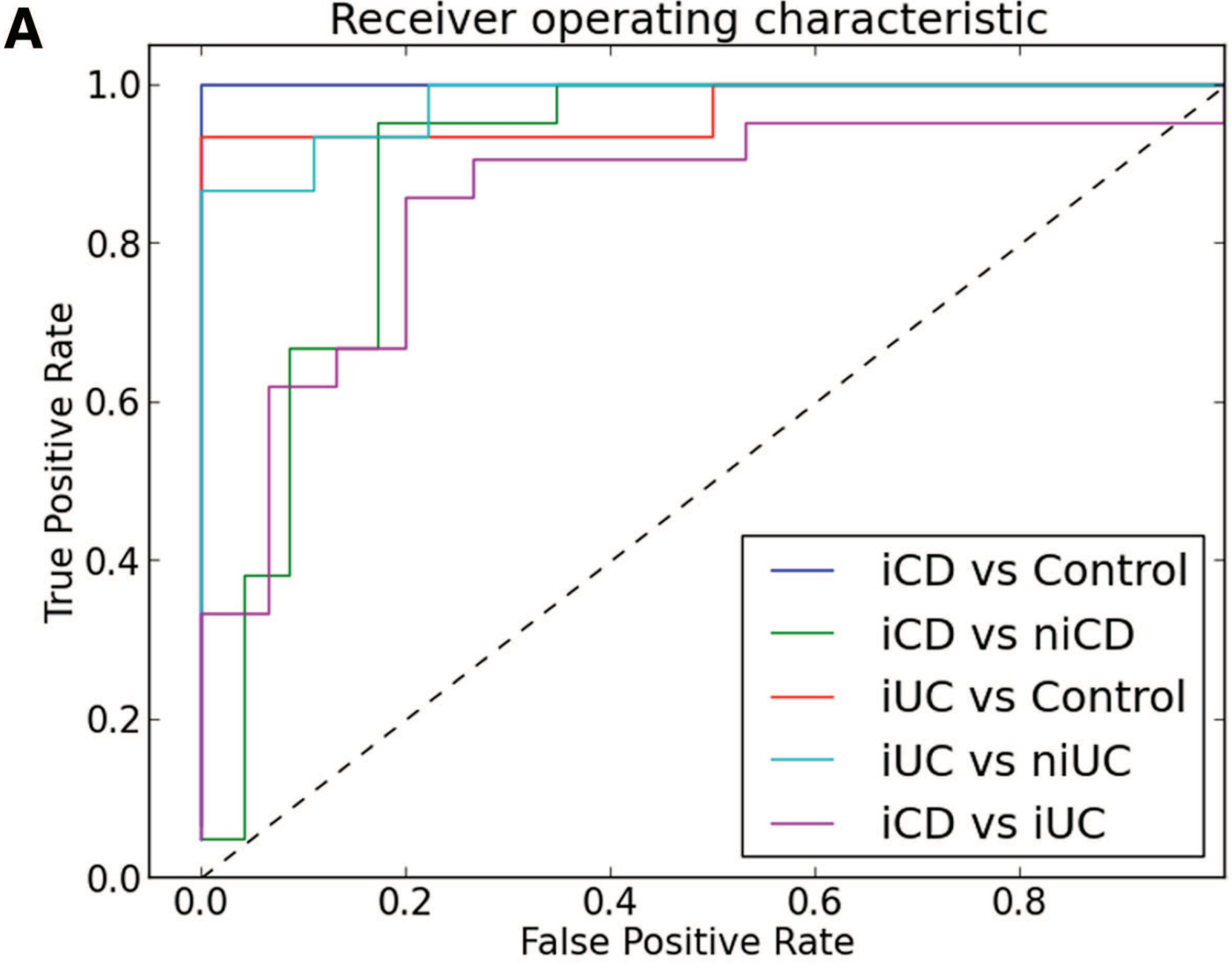

iUC vs Control
log2 Fold Change
30
25
20
15
10
5
0
−5
DEFB1
NPY
POMC
CARD6
CD40LG
RIPK2
IL1R1
CSTA
IRAK3
DEFA3
NOD2
CCR1
CD40
CXCR1
DEFB103B
LY96
CASP1
CCL11
CCL4
SLPI
C1R
CXCR2
CTSG
LTB
CCL19
C3
DMBT1
IL1RN
CXCL9
S100A9
CXCL2
CXCL1

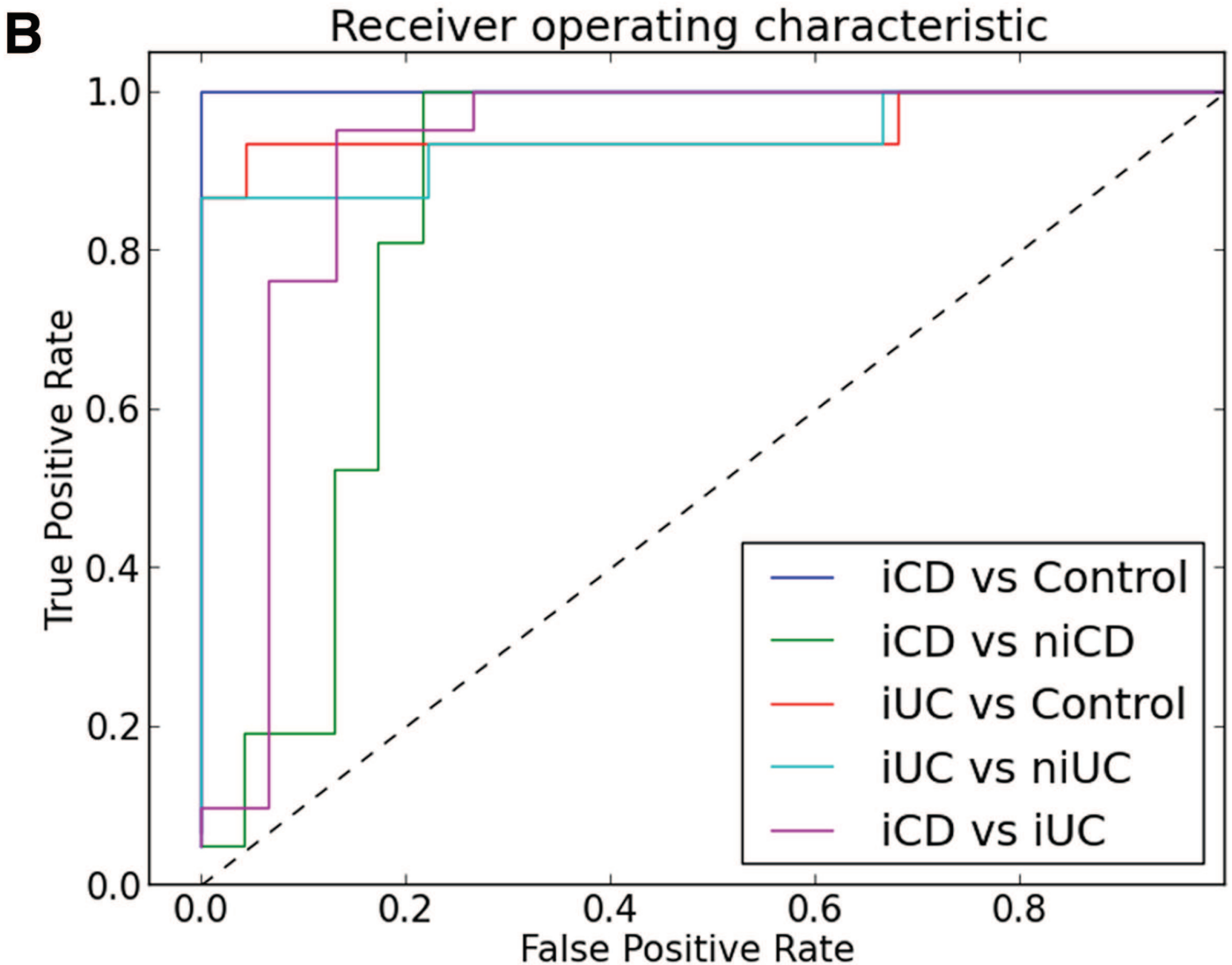

iCD vs niCD
log2 Fold Change
15
10
5
0
−5
NPY
MEFV
RNASE2
RIPK2
CD40
IRAK3
LY96
C1R
CCR1
DEFA3
CARD6
CXCR1
C3
CXCR2
CCL4
CXCL9
IL1RN
S100A9
CXCL2
DMBT1
CXCL1


iUC vs niUC
log2 Fold Change
15
10
5
0
−5
DEFB1
HBA2
NPY
RIPK2
CD40LG
POMC
IL1R1
CARD9
IRAK3
CSTA
LY96
CCR1
NOD2
CXCR1
CD40
CARD6
LTB
CCL4
DEFB103B
CCL11
C1R
CXCR2
CCL19
CASP1
C3
CXCL9
DMBT1
CXCL2
S100A9
CXCL1


A
Color Key
6 8 10 12 14
Value
AC012370.2
AC016747.3
AC019221.4
AC034220.3
CTD-2196E14.5
CTD-2319I12.1
CTD-2540B15.9
LINC00484
LINC00511
RP11-286H14.6
RP11-290F20.3
RP11-300J18.3
RP11-465L10.10
RP11-509E16.1
RP11-542M13.2
RP11-783K16.10
RP11-783K16.5
RP3-395M20.8
S1PR2
SLC37A4
UBAC2-IT1
XXbac-BPG246D15.8
AC006293.3
AC034220.3
AC104699.1
AJ006998.2
AP001058.3
CTB-43E15.4
CTB-61M7.2
H19
IFNG-AS1
RP1-111C20.3
RP11-15F12.1
RP11-336K24.6
RP11-356I2.4
RP11-387H17.4
RP11-536K7.3
RP11-798K3.2
RP11-81H14.2
RP11-98D18.17
AC011193.1
AC051649.12
AC099668.5
HCP5
LINC00511
RP11-464F9.9
RP11-94L15.2
RP3-395M20.8
S1PR2
SLC37A4
XXbac-BPG246D15.8
XXbac-BPG254F23.6
B
iCD
iUC
niCD
niUC
Control
RAB24
LAMB2
CALM3
IFIH1
RASSF1
HLA-DQA2
VWA1
LSP1
LTB
HLA-DQB1
ADCY3
SLC39A7
IL6ST
AHSA2
IKZF1
IFNAR2
TRIP6
LAT
TAP2
C2
PPP1R18
HLA-DQB2
C4B
CXCL2
GMPPB
C21orf33
CCDC85B
GPSM1
UBA7
PRDM1
RASIP1
SIPA1
EFEMP2
CCR10
FAM132A
SLC25A20
HOXA5
ZDHHC11
ZNF768
LRRC66
FBXL20
SLC26A6
SLC22A5
C6orf136
NR5A2
SYNPO
ATP6V0A1
SCARNA17
ESRRA
HLA-DPA1
SEPHS2
SEMA3B
SMPD3
TUFT1
FRYL
SPHK2
DDAH2
DGKD
MST1
CCDC88B
HLA-DPB1
IFITM4P
IER3
HLA-DQA1
SLAMF7
CXCL1
REG4
NDST2
TSPAN33
JAK2
RSPO3
ITIH4
FERMT3
SLAMF1
ITGAL
DOK3
FADS1
TNFRSF4
RASGRP1
GPR65
IL18RAP
TAGAP
HYAL1
NOTCH4
HSPA6
STAT4
TNC
PLAU
COL7A1
GIPC2
APOM
HOXA2
GPR35
HCN3
AMT
THRA
AGFG2
CD6
STAT3
RIPK2
NT5DC2
RASSF5
PCOLCE
BLM
IL1R1
CXCL5
CCL11
SOCS1
TNFRSF6B
FAIM3
GPR18
ANKRD44
FCGR2B
NCR3
IL18R1
STMN3
MXRA8
SP140
CD40
NOD2
PTGFR
CARD11
IKZF3
SYT11
IL15RA
RNF145
NLRP7
PTGIR
POMC
TNFRSF18
IL2RA
CDH3
ENOX1
SNX20
BACH2
CARD9
HLA-DOA
C20orf112
UCN2
IL24
CSF2
CXCR1
CSF3
CXCR2
TNFRSF9
CACNA2D2
BDH2
MYH7B
MUC1
GNAI2
SHISA5
UQCR10
HLA-E
PTRF


A
iCD
iUC
niCD
niUC
Control
ENST00000509204.1 (log2 expression)
LSP1 (log2 expression)
B
ENST00000443574.1 (log2 expression)
HLA−DQB1 (log2 expression)
C
ENST00000563780.1 (log2 expression)
MST1 (log2 expression)
D
ENST00000498745.1 (log2 expression)
TSPAN33 (log2 expression)
E
ENST00000417795.1 (log2 expression)
SLC22A5 (log2 expression)
F
ENST00000442524.1 (log2 expression)
DGKD (log2 expression)


A
Receiver operating characteristic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
iCD vs Control
iCD vs niCD
iUC vs Control
iUC vs niUC
iCD vs iUC


B
Receiver operating characteristic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
iCD vs Control
iCD vs niCD
iUC vs Control
iUC vs niUC
iCD vs iUC

Additional files provided with this submission:

Additional file 1. Table S1, S2, S3 and S4: The top differentially expressed lncRNAs and protein-coding genes for inflamed vs. non-inflamed tissues in CD and UC. Table S1: Top 10 up-down regulated lncRNAs and protein-coding genes in iCD vs niCD. Table S2: Top 10 up-down regulated lncRNAs and protein-coding genes in iUC vs niUC. Table S3: Common differentially expressed genes found between iCD vs control and niCD vs control. Table S4: Common differentially expressed genes between iUC vs control and niUC vs control. **Table S5, S6 and S7**: Table S5: List of oligonucleotides used for validating microarray data using real–time PCR analysis for selected 15 genes. Table S6: Validation of microarray results by Real-time PCR analysis for 8 differentially expressed genes in iCD vs control and iUC vs control. Table S7: Validation of microarray results by Real-time PCR analysis for 5 differentially expressed genes in iCD vs iUC. **Table S8:** 96 differentially expressed lncRNAs found significantly enriched within IBD-loci.**Table S9:** Functional Annotation of differentially expressed lncRNAs based on nearest neighbor approach. **Table S10, S11, S12 and S13:** Weighted correlation network analysis (WGCNA) analysis. Table S10: Gene significance. The co-expression network identified the gene significance for the different clinical parameters. Table S11: Brown module- Immune and inflammatory response: 4216 genes in the module of which 1748 have an Entrez ID (considered in GO analysis).. Table S12 -Green module - Small molecule trans-membrane transport: 2934 genes in the module of which 1210 have an Entrez ID (considered in GO analysis).. Table S13: Red module - Anionic and cationic transport: 2486 genes in the module of which 1173 have an Entrez ID (considered in GO analysis) (625kb)
http://genomemedicine.com/content/supplementary/s13073-015-0162-2-s1.pdf

Additional file 2. Figure S1: The Pearson correlations calculated for the technical replicated (six samples analyzed in duplicates on separate chips: 16_2, 18_3, 27_2, 28_3, 47_3 and 21_2). **Figure S2:** The scatterplot matrices describing the variation explained by the first four principal components for 90 biopsy samples. **Figure S3:** Unsupervised hierarchical clustering of the most dynamic probes (coefficient of variance >0.05) targeting lncRNAs (S3A) and protein-coding genes (S3B) across the samples in different clinical subgroups. **Figure S4:** The log2 ratio and –log10 adj.P-values plotted and represented as volcano plots for the non-inflamed tissues contrasts iCD vs niCD (S4A) and iUC vs niUC (S4B).. **Figure S5:** The expression map of the top 40 differentially expressed lncRNAs and protein-coding genes in iCD vs iUC based on unsupervised hierarchical clustering. **Figure S6:** The expression map of the total differentially expressed lncRNAs and protein-coding genes in iCD vs controls (S6A) and iUC vs controls (S6B) (patients in red, controls in blue) based on unsupervised hierarchical clustering. Figure S7, S8 and S9: Figure S7: Dendrogram of samples and heatmap of clinical parameters. Linear regression model and weighted correlation network analysis (WGCNA) were used to investigate the impact of clinical parameters on disease diagnosis. Figure S8: Receiver operating characteristic (ROC) curve analysis for age (a), sex (b), disease index (c), smoking (d) classification using differentially expressed lncRNAs in all five contrasts. Figure S9: Overlap of differentially expressed genes identified by LIMMA and SVM. **Figure S10 and S11: Figure S10:** Co-expression network is built by hierarchical clustering and Dynamic Tree Cut. Modules are clusters of highly interconnected genes. The “brown”, “green” and “red” modules are enriched for differentially expressed genes between iUC / iCD and Control. Figure S11: Module significance is determined as the average absolute gene significance measure for all genes in a given module (2562kb)
http://genomemedicine.com/content/supplementary/s13073-015-0162-2-s2.pdf